OM nucleic - nucleic search, using sw model

Run on: September 9, 2004, 02:32:16 ; Search time 11655 Seconds
                                              (without alignments)
                                              11610.206 Million cell updates/sec

Title: US-10-088-384A-27
Perfect score: 3122
Sequence: 1 ttcatatattcgacctctc........atgagatagccttgtaagca 3122

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*
               1: gb_ba:*
               2: gb_htg:*
               3: gb_in:*
               4: gb_om:*
               5: gb_ov:*
               6: gb_pat:*
               7: gb_ph:*
               8: gb_pl:*
               9: gb_pr:*
               10: gb_ro:*
               11: gb_sts:*
               12: gb_sy:*
               13: gb_un:*
               14: gb_vi:*
               15: em_ba:*
               16: em_fun:*
               17: em_hum:*
               18: em_in:*
               19: em_mu:*
               20: em_om:*
               21: em_or:*
               22: em_ph:*
               23: em_pat:*
               24: em_ph:*
               25: em_pl:*
               26: em_ro:*
               27: em_sts:*
               28: em_un:*
               29: em_vi:*
               30: em_htg_hum:*
               31: em_htg_inv:*
               32: em_htg_other:*
               33: em_htg_mus:*
               34: em_htg_pln:*
               35: em_htg_rod:*
               36: em_htg_mam:*
               37: em_htg_vrt:*
               38: em_sy:*
               39: em_htgo_hum:*
               40: em_htgo_mus:*
               41: em_htgo_other:*

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Match Length	DB	ID	Description
1		3122	100.0	3122	6	AX099282	AX099282 Sequence
2	c	3122	100.0	81414	8	AB024037	AB024037 Arabidops
3		3122	100.0	110157	8	AF058825	AF058825 Arabidops
4		2029	65.0	74475	8	T7M24	AF077408 Arabidops
5		2029	65.0	196296	8	ATCHRIV10	AL161498 Arabidops
6	c	1973	63.2	68352	8	T15F17	AF262042 Arabidops
7	c	1967.2	63.0	116966	8	AB046433	AB046433 Arabidops
8	c	1937.2	62.0	81902	8	T17A2	AF160183 Arabidops
9	c	1937.2	62.0	195837	8	ATCHRIV21	AL161509 Arabidops
10	c	1925.4	61.7	123460	8	AB073166	AB073166 Arabidops
11	c	1911.6	61.2	14879	8	AC084241	AC084241 Arabidops
12		1911.6	61.2	117055	8	AC073433	AC073433 Arabidops
13		1910.6	61.2	69752	8	TSH22	AF096372 Arabidops
14		1910.6	61.2	159629	8	ATCHRIV9	AL161497 Arabidops
15	c	1905.4	61.0	98536	8	AC074111	AC074111 Arabidops
16		1904.4	61.0	96547	8	AB073154	AB073154 Arabidops
17		1904.4	61.0	104628	8	AB073153	AB073153 Arabidops
18		1904.4	61.0	137179	2	AC109923	AC109923 Arabidops
19		1904	61.0	132522	8	AB062089	AB062089 Arabidops
20		1900.2	60.9	102331	8	AC007120	AC007120 Arabidops
21	c	1899.8	60.9	77287	8	AC026642	AB026642 Arabidops
22	c	1897.6	60.8	114656	8	ATT15B3	AL163975 Arabidops
23		1890	60.5	96684	8	AB046440	AB046440 Arabidops
24	c	1890	60.5	116461	8	AC051625	AC051625 Genomic S
25		1890	60.5	133529	8	AC069553	AC069553 Genomic S
26		1880	60.2	118662	2	AC087569	AC087569 Genomic S
27		1874.2	60.0	81902	8	T17A2	AF160183 Arabidops
28		1874.2	60.0	110826	8	AC007125	AC007125 Arabidops
29	c	1873.2	60.0	90840	8	AB073158	AB073158 Arabidops
30	c	1872.2	60.0	34858	8	AB073165	AB073165 Arabidops
31		1863.2	59.7	195837	8	ATCHRIV21	AL161509 Arabidops
32		1841.8	59.0	85962	8	AB046431	AB046431 Arabidops
33	c	1841.8	59.0	95463	8	F1809	AF296831 Arabidops
34		1831.6	58.7	57000	8	AB046428	AB046428 Arabidops
35		1794.4	57.5	111893	8	AC069555	AC069555 Genomic S
36	c	1794.4	57.5	141430	8	AC069329	AC069329 Genomic S
37		1793.2	57.4	123547	8	AC074109	AC074109 Arabidops
38	c	1744.8	55.9	68639	8	AP001301	AP001301 Arabidops
39		1739	55.7	105420	8	ATCHRIV20	AC006268 Arabidops
40	c	1739	55.7	196517	8	AC007311	AL161508 Arabidops
41		1738	55.7	28670	8	ATF7P3	AC007311 Arabidops
42	c	1698.6	54.4	108227	8	F28J9	AL138663 Arabidops
43	c	1665.2	53.3	115963	8	AX059479	AC007918 Arabidops
44		1638.4	52.5	39104	6	ATF26B15	AX059479 Sequence
45		1638.4	52.5	96988	8		AL138645 Arabidops

ALIGNMENTS

RESULT 1
AX099282
LOCUS   AX099282   3122 bp   DNA   linear   PAT 02-APR-2001
DEFINITION   Sequence 27 from Patent WO0120010.
ACCESSION   AX099282
VERSION   AX099282.1 GI:13538455
KEYWORDS   .
SOURCE   Arabidopsis thaliana (thale cress)
  ORGANISM   Arabidopsis thaliana
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Steiner,A., Mittelsten Scheid,O. and Paszkowski,J.
  TITLE   Transcriptionally silenced plant genes

Pred. No. is the number of results predicted by chance to have a

JOURNAL   Patent: WO 0120010-A 27 22-MAR-2001;
           Syngenta Participations AG (CH)
FEATURES   Location/Qualifiers
  source   1..3122
           /organism="Arabidopsis thaliana"
           /mol_type="unassigned DNA"
           /db_xref="taxon:3702"

ORIGIN

Query Match   100.0%;   Score 3122;   DB 6;   Length 3122;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 TTCATATATTCGACCTCTTCTTCTCATTCTTGCATCCAAAAGACACAACAAGCCGGCCATC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TTCATATATTCGACCTCTTCTTCTCATTCTTGCATCCAAAAGACACAACAAGCCGGCCATC 60

Qy     61 GCTTTCCCTCACAACTTCACTCGACCACCGCTCCTCCACTTACTCGGCTTCATCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCTTTCCCTCACAACTTCACTCGACCACCGCTCCTCCACTTACTCGGCTTCATCG 120

Qy    121 CTCTCATCGCCATCTCTCAACATACTCGACCTCGCGATATCACTCGAGCTCGCCGCTTCT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTCTCATCGCCATCTCTCAACATACTCGACCTCGCGATATCACTCGAGCTCGCCGCTTCT 180

Qy    181 CACCGGCTCTCCATCGTCACCGCCTGCTCCCTTCTCCAAGGAAACAACTCGAGCTCTCC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CACCGGCTCTCCATCGTCACCGCCTGCTCCCTTCTCCAAGGAAACAACTCGAGCTCTCC 240

Qy    241 ATTTCACTCACTCGACCTTCACCACCACCAAGCCGGCTTCTAGCTCTTAACCACT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ATTTCACTCACTCGACCTTCACCACCACCAAGCCGGCTTCTAGCTCTTAACCACT 300

Qy    301 CGACCACCTTCACCACCATCAACAATCAAATCGTTTCTTCCTCCATTAAAGCTTGACATACT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CGACCACCTTCACCACCATCAACAATCAAATCGTTTCTTCCTCCATTAAAGCTTGACATACT 360

Qy    361 CGACCGCTGAACACTTATCACCTTCAAGCTCTTCATCTTCCAACACCGCT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CGACCGCTGAACACTTATCACCTTCAAGCTCTTCATCGTTTCCAACACCGCT 420

Qy    421 GCTCTCATCCCCCAGAAAGCTTGGTCATCACCTCACTCACCTCACTCGATTC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GCTCTCATCCCCCAGAAAGCTTGTCATCACCTCACTCACCTCACTCGATTC 480

Qy    481 AGCAACCAAAACTCGACCTCGTCTTCTTGCCACTCACTCCTCTCCTCTCCTCCACCA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AGCAACCAAAACTCGACCTCGTCTTCTTGCCACTCACTCCTCTCCTCTCCTCCACCA 540

Qy    541 TCTTTCATCATCTCCCTTACTCGACCACCGTGCGTCTCGCTCCGCCCATTGCCATTTAAAAG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TCTTTCATCATCTCCCTTACTCGACCACCGTGCGTCTCGCTCCGCCCATTGCCATTTAAAAG 600

Qy    601 CTCACTCGATTGTCAAAGAAGAAGAGAGTGAAGCTCAACACCGCCACTCGACCGCGTTT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CTCACTCGATTGTCAAAGAAGAAGAGAGTGAAGCTCAACACCGCCACTCGACCGCGTTT 660

Qy    661 CCCTCTCACACATTCAACACTGACCACCGGTGCTACACCGCCACCATCTCCACACCCGA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CCCTCTCACACATTCAACACTGACCACCGGTGCTACACCGCCACCATCTCCACACCCGA 720

Qy    721 CATACACTGACCACCAACAACTCTCAAAGTAAAAAAAAAGAAAAAAAAAGTAAAACCGA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CATACACTGACCACCAACAACTCTCAAAGTAAAAAAAAAGAAAAAAAAAGTAAAACCGA 780

Qy    781 CAGTTGACTTCACTCAACCGGTTTACTCGACCGGTTGGTTTACTCGTGTGTTTTGGTTTT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CAGTTGACTTCACTCAACCGGTTTACTCGACCGGTTGGTTTACTCGTGTGTTTTGGTTTT 840

Qy    841 GCTATTACTAACATATTAACGTTTATCTTTGAGTTTCGTTCTGTTTTTTAGGTTTCATCATG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GCTATTACTAACATATTAACGTTTATCTTTGAGTTTCGTTCTGTTTTTTAGGTTTCATCATG 900

Qy    901 AGTAACTACAGTGTCGAAAATCCTCTATGCACCCTGATTATAATGTGGATGAAGCTAAGTCC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 AGTAACTACAGTGTCGAAAATCCTCTATGCACCCTGATTATAATGTGGATGAAGCTAAGTCC 960

Qy    961 TGGTCCACTAGACCGGAGTGAGCAACATGTTACGAGAGCTAGAGAGATGAATTTGAA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TGGTCCACTAGACCGGAGTGAGCAACATGTTACGAGAGCTAGAGAGATGAATTTGAA 1020

Qy   1021 CGCTCTCGAGCTCGACGTAATCAAAGAGAGCTGAAATCGCTAGAGAGAAAGAGAGGGCGATG 1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CGCTCTCGAGCTCGACGTAATCAAAGAGAGCTGAAATCGCTAGAGAGAAAGAGAGGGCGATG 1080

Qy   1081 TCGAGTAGATATGAGCTGATTGATGAGGATATCAAAACTGAGTATGAGCCAGAGTCATCG 1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TCGAGTAGATATGAGCTGATTGATGAGGATATCAAAACTGAGTATGAGCCAGAGTCATCG 1140

Qy   1141 CGCAAGGAGGACGAAGCTACTGAACAAATCCGACGAGGTTACAGTGGAGAGTATATCAGA 1200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CGCAAGGAGGACGAAGCTACTGAACAAATCCGACGAGGTTACAGTGGAGAGTATATCAGA 1200

Qy   1201 TTCTTTGAGATGAATGACTTCTCGGGGAACGAGCTATCCTGATATGAGACTTTAGCCCAG 1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 TTCTTTGAGATGAATGACTTCTCGGGGAACGAGCTATCCTGATATGAGACTTTAGCCCAG 1260

Qy   1261 TTGGGGTTACTGAGGAGACGTGCAGCATCTGTTCGAGAAGTGTCATCTGATAAGAGGAGA 1320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 TTGGGGTTACTGAGGAGACGTGCAGCATCTGTTCGAGAAGTGTCATCTGATAAGAGGAGA 1320

Qy   1321 CAATCGAGTTCTTCCACACTGCAAGTGGAAATGTATGAGGGACTCACAGACTTTGAGC 1380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CAATCGAGTTCTTCCACACTGCAAGTGGAAATGTATGAGGGACTCACAGACTTTGAGC 1380

Qy   1381 TGGATACCATGGGGTTAGGCTTCTTGACGTTCTTAGTGGATGAACAGCGGTACCAGATTT 1440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TGGATACCATGGGGTTAGGCTTCTTGACGTTCTTAGTGGATGAACAGCGGTACCAGATTT 1440

Qy   1441 AGATCAAGAAATTCGAAGAACTGTTTGGTTTCCCTAGTGAAAGGGAACCAACCCCAGGT 1500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 AGATCAAGAAATTCGAAGAACTGTTTGGTTTCCCTAGTGAAAGGGAACCAACCCCAGGT 1500

Qy   1501 TTGCAGGGAAGAGCTTAAGGATTGTGGGCTACTATTGGGAACAATCTACCGCTAAACT 1560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TTGACAGGGAAGAGCTTAAGGATTGTGTGGGCTACTATTGGGAACAATCTACCGCTAAACT 1560

Qy   1561 CGACCCGGTCCAAGAGCAACAACCAAATCCGAGAGTCCTGTGATTCGCTACTTTCAGCGCTCG 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CGACCCGGTCCAAGAGCAACAACCAAATCCGAGAGTCCTGTGATTCGCTACTTTCAGCGCTCG 1620

Qy   1621 TTGCCAATGTTTTTACTCCAGGGACACCGGTTGTCTAACACGACACATGAAGA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 TTGCCAATGTTTTTTACTCCAGGGACACCGGTTGTCTAACACGACACATGAAGA 1680

Qy   1681 TGATAGATTCAGCGCTTATAGGGATTCTCCGCCTTACAAAAGGAAAGAAATGTCCTGAGAG 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TGATAGATTCAGCGCTTATAGGGATTCTCCGCCTTACAAAAGGAAAGAAATGTCCTGAGAG 1740

Qy   1741 GAGATCTTAACGACTCACCAGTAATGCCTCGTGTTGATCCATCTGTGTGGGTACATGA 1800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 GAGATCTTAACGACTCACCAGTAATGCCTCGTGTTGATCCATCTGTGTGGGTACATGA 1800

Qy   1801 AGTGGCGCTTGACAAACGGCAAGAAGGTAAGACAGGACACTATGCGTGGTGGCGTTG 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 AGTGGCGCTGACAAACGGCAAGAAGGTAAGACAGGACACTATGCGTGGTGGCGTTG 1860

Qy   1861 TGATGGCCAATTCTGAAAGTTCTGAAAGTTGTGGAGTTCCGCTCAAGGAAGTAGGGTTAGCACCGGAA 1920
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 TGACGGCCAATTCTGAAAGTTCTGAAAGTTGTGGAGTTCCGCTCAAGGAAGTAGGGTTAGCACCGGAA 1920

Qy   1921 TGATGGACTTGGATCACTTGCGCCGATGTGAGTTCTCGAGTTGACATGGTTGGCGACT 1980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 TGATGGACTTGGATCACTTGCGCCGATGTGAGTTCTCGAGTTGACATGGTTGGCGACT 1980

Qy   1981 TTCACCGCTACAGGTTCGAGCATTCATCGATTAGAATCGCCAACATTCTTTTCCCCTGCA 2040
```

```
Db   3061 AGTGTTTCATTTAGGATTGTTGCATATGCATAGGGGATAATGATGAGATAGCCTTGTAAG 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3121 CA 3122
          ||
Db   3121 CA 3122
```

RESULT 2
AB024037/c
LOCUS   AB024037   81414 bp   DNA   linear   PLN 09-AUG-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MSK10.
ACCESSION   AB024037
VERSION   AB024037.1 GI:4519196
KEYWORDS   .
SOURCE   Arabidopsis thaliana (thale cress)
  ORGANISM Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (sites)
  AUTHORS   Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E., Kotani,H.
            and Tabata,S.
  TITLE   Structural analysis of Arabidopsis thaliana chromosome 5. X.
            Sequence features of the regions of 3,076,755 bp covered by sixty
            P1 and TAC clones
  JOURNAL   DNA Res. 7 (1), 31-63 (2000)
  MEDLINE   20181125
   PUBMED   10718197
REFERENCE   2 (bases 1 to 81414)
  AUTHORS   Nakamura,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (24-FEB-1999) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
            Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT   Address for correspondence: kaos@kazusa.or.jp
            For the latest information on annotation of this clone, please see
            http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MSK10
            Genes with similarity to proteins in the databases are described in
            'product' or 'note' qualifiers. Genes that have no significant
            protein similarity are described as 'unknown protein'.
            The software programs used to predict genes include: Grail
            (Informatics Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/Grail-1.3/),
            GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
            NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
            Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
            SplicePredictor (Volker Brendel, Stanford University,
            http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi).
            Genes encoding tRNAs are predicted by tRNAscan-SE
            (Sean Eddy, Washington University School of Medicine, St. Louis,
            http://genome.wustl.edu/eddy/tRNAscan-SE/).
            This sequence may not be the entire insert of this clone. It may be
            shorter because we remove overlaps between neighboring submissions.
            The 5' clone is MGG23 and the 3' clone is T13C12.
FEATURES   Location/Qualifiers
     source   1..81414
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="MSK10"
     exon   complement(1..192)
                     /clone_lib="Mitsui P1"
                     /gene_id:MGG23.7"
                     /product="retroelement pol polyprotein-like"
                     /note="CDS is reported in Acc# AB028613
     exon   complement(244..2797)
                     /number=2
                     /evidence=not_experimental
                     /product="retroelement pol polyprotein-like"
                     /note="CDS is reported in Acc# AB028613

```
Db   1981 TTCACCGCTACAGGTTCGAGCATTCATCGATTAGAATCGCCAACATTCTTTCCCCTGCA 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2041 TTTACGCTACTAGGATTCTCGAGGCAGGAACATTGACTTCAAGCCTCGGCTTGAAGATC 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 TTTACGCTACTAGGATTCTCGAGGCAGGAACATTGACTTCAAGCCTCGGCTTGAAGATC 2100
Qy   2101 TTTATTTCGAGGGGCAGTCCGCCAACTGAGGAGATTAGTCACACCGAGGAGCTACAATAG 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 TTTATTTCGAGGGGCAGTCCGCCAACTGAGGAGATTAGTCACACCGAGGAGCTACAATAG 2160
Qy   2161 AAGATGTTGATGAGACATATGATATAGATGAGGCGGAGTTGACACGAGCATGTATCATT 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 AAGATGTTGATGAGACATATGATATAGATGAGGCGGAGTTTGACACGAGCATGTATCATT 2220
Qy   2221 TCAGTGAGCATATACCTCCAGCGAGCGAGAAAGCAAGAGTTTGAGCGAAGCTCACAGGAACA 2280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 TCAGTGAGCATATACCTCCAGCGAGCGAGAAAGCAAGAGTTTGAGCGAAGCTCACAGGAACA 2280
Qy   2281 ACAGCAAGCTGCAGAAGTGGTGCAAGAAACAGGATAAGTTACTCGCCAAGTGCCTCAGGG 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 ACAGCAAGCTGCAGAAGTGGTGCAAGAAACAGGATAAGTTACTCGCCAAGTGCCTCAGGG 2340
Qy   2341 CTATCAAGTTTCTGAAGGACAAGATCAGCTGCTCTCTTCCACTACAACTATTCCGCAAT 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CTATCAAGTTTCTGAAGGACAAGATCAGCTGCTCTCTTCCACTACAACTATTCCGCAAT 2400
Qy   2401 GACAGCTCCCTCAGGACATGCCTTCGAGGAGATATGACGCGCCCGAGCCTAGAGAGCAGA 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GACAGCTCCCTCAGGACATGCCTTCGAGGAGATATGACGCGCCCGAGCCTAGAGAGCAGA 2460
Qy   2461 AGATTCTGCATGCCCTGCCTCGGAGGCATTCATCATTCGAGCCTCGTGAATCTAGGAAGAATA 2520
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 AGATTCTGCATGCCCTGCCTCGGAGGCATTCATCATTCGAGCCTCGTGAATCTAGGAAGAATA 2520
Qy   2521 GGAGAACGACACTCACTCGATCTAGCAGCAGGAGCAGACACTTCTCGCAGTCTCGTAGTT 2580
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 GGAGAGAACGACACTCACTCGATCTAGCAGCAGGAGCAGACACTTCTCGCAGTCTCGTAGTT 2580
Qy   2581 TACGGCGACCGCGGTGCTGGCCCAATAGAAGAGAGAGGTCGAGTATCCTCAGAGCGGTG 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 TACGGCGACCGCGGTGCTGGCCCAATAGAAGAGAGAGGTCGAGTATCCTCAGAGCGGTG 2640
Qy   2641 CTGGCCCCACAGAGCTGATGAGATCGAGTACCCACATGCGAGCTGGAGCTGATACGGAACATG 2700
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 CTGGCCCCACAGAGCTGATGAGATCGAGTACCCACATGCTGAGCTGATACGGAACATG 2700
Qy   2701 GCGGTTCGTCTATGGCCTTGGGAGCAATCACAGGCAGCCATTGACTACCAACTTCGTTCAT 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 GCGGTTCGTCTATGGCCTTGGGAGCAATCACAGGCAGCCATTGACTACCAACTTCGTTCAT 2760
Qy   2761 TATTCGACTGAGTAAGCGCCTCACTTCACCATTATATATCATCTCTTGTGATTGT 2820
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 TATTCGACTGAGTAAGCGCCTCACTTCACCATTATATATCATCTCTTTGATTGT 2820
Qy   2821 TCTTTATTTTGTTTCAGTGATTGGATTTGTCCTGAGTACTCCTCTTCCAAGTTTATTCACA 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 TCTTTATTTTGTTTCAGTGATTGGATTTGTCCTGAGTACTCCTCTTCCAAGTTTATTCACA 2880
Qy   2881 CAGTGGACTGTGTGATTTAAGTTTGGGGGAGGGCTCAGGAAGTATGTTGCATTGTATATA 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 CAGTGGACTGTGTGATTTAAGTTTGGGGGAGGGCTCAGGAAGTATGTTGCATTGTATATA 2940
Qy   2941 TTTTTAAGTCTGCATTCATCTAAGGCATAGAAAAACCAAAAAAAAATTAAAAATTTCAGA 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 TTTTTAAGTCTGCATTCATCTAAGGCATAGAAAAACCAAAAAAAAATTAAAAATTTCAGA 3000
Qy   3001 AAATGATTTCACAAAAAAAGAGTGTTCATGTAGTTGCATTACATTTAGGATCAAGTCAG 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 AAATGATTTCACAAAAAAAGAGTGTTCATGTAGTTGCATTACATTTAGGATCAAGTCAG 3060
Qy   3061 AGTGTTTCATTTAGGATTGTTGCATATGCATAGGGGATAATGATGAGATAGCCTTGTAAG 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
CDS          gene_id:MGG23.7"
             /number=1
             /evidence=not_experimental
             complement(408l..4617)
             /note="emb|CAB77928.1
             gene_id:MSK10.2
             similar to unknown protein"
             /codon_start=1
             /evidence=not_experimental
             /protein_id="BAA97394.1"
             /db_xref="GI:8843868"
             /translation="MIRIALHAYCISELFSGDLELFASKLEETSQNQKTYSTPRTSDF
             GAILPIHSTTRSSNIASSGHSMTLLDPLVESHFATPPDYTRPFTLPRYSTPWSSIITHH
             HHHISTGHSITSSQSTQIRTQPDKLSTRKRRGEDKVFGSGLDLHRITKPISAHYHSMG
             RAIRLLARLLSFYFWYK"
CDS          join(4695..5288,5361..5486,5645..5724,5835..5960,
             6033..6137,6227..6328,6399..6458,6542..6629,6876..7058)
             /note="emb|CAB81580.1
             gene_id:MSK10.3
             strong similarity to unknown protein"
             /pseudo
             /codon_start=1
             /evidence=not_experimental
             /protein_id="BAA97395.1"
             /db_xref="GI:8843869"
CDS          7427..8119
             /note="contains similarity to mutator-like transposase
             gene_id:MSK10.4"
             /pseudo
             /codon_start=1
             /evidence=not_experimental
             8890..9582
             /note="contains similarity to mutator-like transposase
             gene_id:MSK10.5"
             /pseudo
             /codon_start=1
             /evidence=not_experimental
             /protein_id="BAA97395.1"
             /db_xref="GI:8843869"
             /translation="MPNPNIPDDLLARIAKKMADKCWWYLGPMLKSGPRGRDIVRPD
             VLKDANIFSMCDDPDDFYAAGHDPNDINSEDHYRPFFKRCLQAGNRTTIYHEGLRIVT
             HESDIQGAILHLERIAPRNAAATFTCAILYICAGNAHMGGVYLRLFGSNHYALESEEA
             RDICEEVLEDIKKYGNTLKYTYANGFSYPECGDVSTPKCAEMCYMRSRLFNNLCNQCY
             IWLCAKRISQIL"
CDS          join(10112..10220,10432..10664)
             /note="gene_id:MSK10.6
             pir||T02599
             similar to unknown protein"
             /codon_start=1
             /evidence=not_experimental
             /protein_id="BAA97396.1"
             /db_xref="GI:8843870"
             /translation="MKPYCSYYMQYKQTSSQPSVKKRNEYFIDVHGAIMRGIQLPNPI
             SILVNYLPVRQTPIETLRNLVYSCDINTLAIDEFMTWPSLLQPESTRPYFELLVERG
             SLKVFAKRATS"
             complement(11072..15536)
             /note="gb|AAF24957.1
             gene_id:MSK10.7
             strong similarity to unknown protein"
             /pseudo
             /codon_start=1
             /evidence=not_experimental
             join(17864..18351,18409..18718,18998..19184,19258..19718,
             20145..20803,20889..20997,21270..21941)
             /note="gene_id:MSK10.8"
             /codon_start=1
             /evidence=not_experimental
             /product="retroelement pol polyprotein-like"
             /protein_id="BAA97397.1"
             /db_xref="GI:8843871"
             /translation="MSEDQAREYPPRLYQEGDSNLEGKGINHNIKLGEFPAIRESIGK
             EIWAELKETELRLIAKLVDSYFLWSGKTVHYLLCRQLRILKELWCIVAGKPIRFGLN
             EFHHITGLNIDELPTEKFVPAADYKAFFSELIVPTGEGPNLDELRQGLVTCRAWPTEK
             RWIPFESAKRVFDDKAMKTYPWGRAAYEALVVSIKLLRPIGKTYTVSGLVFVLQAWA
             YESIITIAERFGNAVNADEILLLRWGGNRTRATIESVIAEDIKANGGBLRVRTMVEKA
             ELSELFHFWPDQLEDPALDNLIQDLHGVKLIVRGYWDVKTSEKKKKLLKDKSREKKKKK
```

```
             KNKKKEVAVBSDEDEDVRSREQDGQGDEQSLVDVVSNLISPLNSRFDTVDSIKEMS
             RLDVIGCQIESSVEAKFEGRFGPSIENDVKQIKEQLKAIADSKSSSYIRDMFLAKPQPQ
             TQEDNPKAQTQKTPDVPSKKITNNQSTAPSPPSKHADVGVSRKLEABFDKASIVSDFD
             FVETVSPLNRRLRSRKPPAPNVPKQKKEKTLNELIQKPPARRGRGRKPLEQPKKVPPT
             APKIRIKKPKPSEETKEKAEEAKEABEDSDVVDVTDKVVSEYNEMLPESDEDEETE
             RLKSVKEIRLKTVKLAPDGSSLINLBDMPTHTFPVIGDNGLSCMRKGCLPSRAIYDPL
             APVDPAKLQKLKAYLTGPESKHYKTYGFDDVLFYSVIIAERKDWPDKKHGWLGDDHI
             AAYVKVLTTRLLRDFTPYYSERVMILDPWFISMWSRDYAQWKMKPKRVSFKGSTYETW
             VNGTGGEDPTNKKWIPDVDHLYTILQTGGNWATIHVDLPRGHVDCYDCIVGCHTKES
             DGKILEHCRPFTRMIPQIKSEIIPPEVRVPQYDQFSFQRRDLKKVPQNTITGDCGVYT
             LKILECLLLGVSFEGIPDSNIQGLRVRMATFIYDELPNRMSIRFYET"
             complement(join(23427..23636,23723..23893,23970..24017))
             /note="gene_id:T0166
             pir||T0166
             similar to unknown protein"
             /codon_start=1
             /evidence=not_experimental
             /protein_id="BAA97398.1"
             /db_xref="GI:8843872"
             /translation="MMAQRSLFHGESSVDEELNRMEIRMRTRYDKNRRRDKGVPIECD
             CNAKVVVATSLDPVTTGKLFFSCPHEISDGPQGQCGFKRWTVALCDEFDMIKEETTE
             MKKDLEAANKRVESQTEKIFlMEKKFETLEKKYESINKYL"
             complement(24565..24837)
             /note="gene_id:MSK10.10
             unknown protein"
             /codon_start=1
             /evidence=not_experimental
             /protein_id="BAA97399.1"
             /db_xref="GI:8843873"
             /translation="MTMIIKARQKQQSLIINFHFWFSLSLLITSSPSSFTIACNASSS
             TETSDLAALNSLSKPANFSQLSTTTSYVPSSTHLNKWSCIFNHIRN"
             complement(join(25379..25591,25667..25897))
             /note="gene_id:MSK10.11"
             /codon_start=1
             /evidence=not_experimental
             /product="En/Spm-like transposon protein-like"
             /protein_id="BAA97400.1"
             /db_xref="GI:8843874"
             /translation="MVPHIDGQPLRDVEAVLTAPDRGLGTWVDKTTVTWFKPNHQLAV
             AIGDILQRNFHHAWPTYRMIPVVFKARWFFLLTRMHTWDSRHTVTWTQFNLVARTLY
             RDNMRVLKRGWARGLDKPRWMTTTVWNDLVDMFTEFDPDDVGFMS"
             join(26245..26379,26472..26736,26817..26919,27018..27261)
             /note="gene_id:MSK10.12"
             /codon_start=1
             /evidence=not_experimental
             /product="replication protein A1-like"
             /protein_id="BAA97401.1"
             /db_xref="GI:8843875"
             /translation="MGYDGLRDLNPYITCWTICVMLRKYVSFITDNSFKLRVFVDE
             WGIQIEAIIGSRRSQRYNMLBENQWKEIRTFGVIENCGFVRTTHKYMIEFMNATEL
             NHSDPRENVLFNRFTPFDYIIEDTAHTENLVDLVGALVEVGCLRNITEDGWEMDGFRL
             TFESMDKYKERLKCEALGELAVDLHRKFRSFRNDNIVIDNIYVALTWWKIFRYHPCPN
             HVKIQNHARISDVTVDPNIEEVQELKMVLE"
             complement(join(27691..28087,28296..29179,29384..30523))
             /note="gene_id:MSK10.13"
             /codon_start=1
             /evidence=not_experimental
             /product="mutator-like transposase"
             /protein_id="BAA97402.1"
             /db_xref="GI:8843876"
             /translation="MAGPATVIVLSGLWVKKQRFVFNPDERGCRVVQLRAERAVYDSL
             LKMVVDVGVNQSTHQLRLSYMFSNKTLKTMPFDTPPVYISNDRQLQCYLGLTKFEQL
             RLCVEFTAKDSSDISRKDGSRGRGSKRAIASDSRAESRFEVLGWGFEGVCYDYQDDEA
```

```
Query Match         100.0%;  Score 3122;  DB 8;  Length 81414;
Best Local Similarity  100.0%;  Pred. No. 0;
Matches 3122;  Mismatches 0;  Conservative 0;  Indels 0;  Gaps 0;

Qy      1 TTCATATATTCGACCTCTTCCATTCTTGCCATCCAAAAGACACAACAAGCCGGCCATC  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55627 TTCATATATTCGACCTCTTCCATTCTTGCCATCCAAAAGACACAACAAGCCGGCCATC  55568

Qy     61 GCTTCCCTCACAACTCTCACTCGACCACCGCCCGCTTCTCCTCACTTACTCGCTTCATCG  120
```

```
Db  55567  GCTTTCCCTCACAACTTCACTCGACGACCACCGCCGCCCGGTCTCGCTCTCACTTACTCCGGTCGCGGTCATCG  55508

Qy    121  CTCTCATCGCCATCTCAACATACTCGACCTCGCGATATCGACCTCGCGATATCGACCTCGAGCTCTCC  180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55507  CTCTCATCGCCATCTCAACATACTCGACCTCGCGATATCGACCTCGCGATATCGACCTCGAGCTCTCT  55448

Qy    181  CACCGGCCTCTCCATGCGACCTGCTCCTCCTCCCAAGGAACAACTCGAGCTCTCC  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55447  CACCGCCTCTCCATGTCAACTGCTCCTCCTCCCAAGGAAACAACTCGAGCTCTCC  55388

Qy    241  ATTTCACTCGACCTCTACCACCAAGCCGGCTTCACCACTTCTTAGCTCTTAACCACT  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55387  ATTTCACTCGACCTCTACCACCAAGCCGGCTTCACCACTTCTTAGCTCTTAACCACT  55328

Qy    301  CGACCACCTTCACCATCAACAATCAAATGTTTCTCCTCCATTAAAGCTTGACATACT  360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55327  CGACCACCTTCACCATCAACAATCAAATGTTTCTCCTCCATTAAAGCTTGACATACT  55268

Qy    361  CGACCGCTGAACACTTATCACCTTCAAGCTCCTCATCGTTTCAACACCGCT  420
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55267  CGACCGCTGAACACTTATCACCTTCAAGCTCCTCATCGTTTCAACACCGCT  55208

Qy    421  GCTCTCATCCCCACGAAAGCTCGTCTCTCTTCACTCACCTCTCCTCCACCA  480
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55207  GCTCTCATCCCCACGAAAGCTCGTCTCTCTTCACTCACCTCTCCTCCACCA  55148

Qy    481  AGCAACCAAACTCGACCTCGTCTCTTCGCCACTCATAGTCACTCGATCTCTCCTCCACCA  540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55147  AGCAACCAAACTCGACCTCGTCTCTTCGCCACTCATAGTCACTCGATCTCTCCTCCACCA  55088

Qy    541  TCTTCATCATCTCCCTTACTCGACCGGTTCGCTCGCCACCACCGCCACCGGTTCAAAAG  600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55087  TCTTCATCATCTCCCTTACTCGACCGGTTCGCTCGCGTGCACCACCGCCACCGGTTCAAAAG  55028

Qy    601  CTCACTCGATTGTCAAAGAAGAAGTGAAGCTCAACCACCGCCACTCGACCGCGTTT  660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55027  CTCACTCGATTGTCAAAGAAGAAGTGAAGCTCAACCACCGCCACTCGACCGCGTTT  54968

Qy    661  CCCTCTACACATTCGACCTGCTACCATCTCCACCACCGCTCTTGTTCAC  720
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db  54967  CCCTCTACACATTCGACCTGCTACCATCTCCACCACCCGCTCTTGTTCAC  54908

Qy    721  CATACACACTCGACCAACACTCTCAAAGTAAAAGAAAAAAGAAAGTCAAAACCGA  780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54907  CATACACACTCGACCAACAACTCTCAAAGTAAAAGAAAAAAGAAAGTCAAAACCGA  54848

Qy    781  CAGTTTCACTCAACCGGTTTACTCGACCGTACGCTGGTTTAGATTGTGTTTTTGGTTTT  840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54847  CAGTTTCACTCAACCGGTTTACTCGACCGTACGCTGGTTTAGATTGTGTTTTTGGTTTT  54788

Qy    841  GCTATTACTCAACATATTAACGTTTATCTTGAGTTCGTCTCGTTTTAGGTTTCATCATG  900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54787  GCTATTACTCAACATATTAACGTTTATCTTGAGTTCGTCTCGTTTTAGGTTTCATCATG  54728

Qy    901  AGTAACTACAGTGGAAAATCCTGATTATAAATGTGGATGAAGCTAAGTCC  960
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54727  AGTAACTACAGTGGAAAATCCTGATTATAAATGTGGATGAAGCTAAGTCC  54668

Qy    961  TGGTCCACTAGACCGGAGTGAGAGCAACATGTTTACGAGCTATAGGGATGAATTTGAA  1020
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54667  TGGTCCACTAGACCGGAGTGAGAGCAACATGTTTACGAGCTATAGGGATGAATTTGAA  54608

Qy   1021  CGCTCTGCAGCTCGACGTAATCAAAGAAGAGCTGAAATCGCTAGAGAAGAGGGGCATG  1080
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54607  CGCTCTGCAGCTCGACGTAATCAAAGAAGAGCTGAAATCGCTAGAGAAGAGGGGCATG  54548

Qy   1081  TCGAGTAGTAGACATATGAGCTGACGTAATCGTCAAAACTGAGTATGAGCCAGAGTCATGG  1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54547  TCGAGTAGTAGACATATGAGCTGACGTAATCGTCAAAACTGAGTATGAGCCAGAGTCATGG  54488

Qy   1141  CGCAAGGAGACGAAGCTACTGAACAAATCCGACGAGAGTTACAGTGGAGGAGTATATCAGA  1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||


Db  54487  CGCAAGGAGGAGACGAAGCTACTGAACAAATCCGACGAGAGTTACAGTGGAGGAGTATATCAGA  54428

Qy   1201  TTCTTTGAGATGAATGACTTCTGGGGACGAGCAGGTATCCTGATATGAGACTTTAGCCCAG  1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54427  TTCTTTGAGATGAATGACTTCTGGGGACGAGCAGGTATCCTGATATGAGACTTTAGCCCAG  54368

Qy   1261  TTGGGGTTACTGGAGACGTGCAGCATCTGTTGTCGAGAAGTGTCATCTGATAAGGAGGAGA  1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54367  TTGGGGTTACTGGAGGACGTGCAGCATCTGTTGTCGAGAAGTGTCATCTGATAAGGAGGAGA  54308

Qy   1321  CAATCGAGTTCTTTCCACACTGCAAGTGAAATGTATGAGGGACTCACGACTTTGAGC  1380
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54307  CAATCGAGTTCTTTCCACACTGCAAGTGGAAATGTATGAGGGACTCACGACTTTGAGC  54248

Qy   1381  TGGATACCCATGGGGTTAGGCTTCTTGACGTTCTTAGTGGATGAACAGCGGTACCAGATTT  1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54247  TGGATACCCATGGGGTTAGGCTTCTTGACGTTCTTAGTGGATGAACAGCGGTACCAGATTT  54188

Qy   1441  AGATCAAGAAATTGGAAGAACTGTTTGTTCCCTAGTGGAAAGGGAACCACCCCCAGGT  1500
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54187  AGATCAAGAAATTGGAAGAACTGTTTGTTCCCTAGTGGAAAGGGAACCACCCCCAGGT  54128

Qy   1501  TTGACAGGGAGAGCTTAAGGATTTGTGGGCTACTATTGGGAACAATCTACCGCTAAACT  1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54127  TTGACAGGGAGAGCTTAAGGATTTGTGGGCTACTATTGGGAACAATCTACCGCTAAACT  54068

Qy   1561  CGACCGCGGTCCAAGAGCAACCAAATCCGAGTCCTGTGATTCGCTACTTTCAGCGCTCGG  1620
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54067  CGACCGCGGTCCAAGAGCAACCAAATCCGAGTCCTGTGATTCGCTACTTTCAGCGCTCGG  54008

Qy   1621  TTGCCAATGTTTTTACTCCAGGAGTCTACAGGCACCGTGTCTAACACAGACATGAAGA  1680
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54007  TTGCCAATGTTTTTACTCCAGGAGTCTACAGGCACCGTGTCTAACACAGACATGAAGA  53948

Qy   1681  TGATAGATTCAGCGCTTATAGGGATTCTCCGCCTTACAAAAGGAAAGAAATGTCCTGAGAG  1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53947  TGATAGATTCAGCGCTTATAGGGATTCTCCGCCTTACAAAAGGAAAGAAATGTCCTGAGAG  53888

Qy   1741  GAGATCTTAACGACTCACCAGTAATGCGCTCGTTGATCCATCGTGTGTGGTACATGA  1800
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53887  GAGATCTTAACGACTCACCAGTAATGCGCTCGTTGATCCATCGTGTGTGGTACATGA  53828

Qy   1801  AGTGGCGCTGACAAACGGCAAGAAGGTAAGACGCACTATGCGTGGCGGTGCGTTG  1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53827  AGTGGGCGCTGACAAACGGCAAGAAGGTAAGACGCACTATGCGTGGCGGTGCGTTG  53768

Qy   1861  TGACGCCAATTCTGAAAGTTGTGGAGTTCGCGGTCTCAAGGAAGTAGGGTTAGCACCGAGAA  1920
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53767  TGACGCCAATTCTGAAAGTTGTGGAGTTCGCGGTCTCAAGGAAGTAGGGTTAGCACCGAGAA  53708

Qy   1921  TGATGGACTTGGATCACTTGCGCCCATGTGAGTCTCTGAGTTCTCTCGAGTTTGACATGGTTGGCGACT  1980
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53707  TGATGGACTTGGATCACTTGCGCCCATGTGAGTCTCTGAGTTCTCTCGAGTTTGACATGGTTGGCGACT  53648

Qy   1981  TTCACCGCTACAGGTTCGAGCATTCATCGATTAGAATCGCCAACATTCTTTCCCCTGCA  2040
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53647  TTCACCGCTACAGGTTCGAGCATTCATCGATTAGAATCGCCAACATTCTTTCCCCTGCA  53588

Qy   2041  TTTACGCTACTAGGATTCTCGAGGGCAGGAACATTGACTTCAAGCCTGCGCTTGAAGATC  2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53587  TTTACGCTACTAGGATTCTCGAGGGCAGGAACATTGACTTCAAGCCTGCGCTTGAAGATC  53528

Qy   2101  TTTATTTGGAGGGGCAGTCGCCAACTAGTAGTCGCTAGAGGAGAAGGAGCTACAATAG  2160
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53527  TTTATTTGGAGGGGCAGTCGCCAACTAGTAGTCGCTAGAGGAGAAGGAGCTACAATAG  53468

Qy   2161  AAGAGTGTTGATGAGACATATGATATAGATGAGGCGGAGTTTGACGAGCATGTATCATT  2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53467  AAGAGTGTTGATGAGACATATGATATAGATGAGGCGGAGTTTGACGAGCATGTATCATT  53408

Qy   2221  TCGAGTGAGCATATACCTTCCAGCGAGGAGAAAAGCAAGAGTTTGAGCGAGCTCACAGAACA  2280
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53407  TCAGTGAGCATATACCTTCCAGCGAGGAGAAAAGCAAGAGTTTGAGCGAGCTCACAGAACA  53348
```

```
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 110157)
  AUTHORS   Washington University Genome Sequencing Center.
  TITLE     The A. thaliana Genome Sequencing Project
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 110157)
  AUTHORS   Dante,M.
  TITLE     The sequence of A. thaliana F7N22
  JOURNAL   Unpublished (1998)
REFERENCE   3  (bases 1 to 110157)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-APR-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     Submitted by:
            Genome Sequencing Center
            Department of Genetics, Washington University,
            St. Louis, MO 63108, USA
            e-mail: rwilson@watson.wustl.edu

            MAPPING: Clones were assigned to the YAC map by hybridization by
            M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted
            by M. Marra, WashU, to pick the best candidates for sequencing.

            NOTICE: This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone

            NEIGHBORING COSMID INFORMATION:

            The actual start of this clone is at base position 1 of F7N22;
            actual end is at 110157 of F7N22.

            NOTES:

            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES             Location/Qualifiers
     source          1..110157
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="IV"
                     /map="unknown"
                     /clone="F7N22"
     misc_feature    3032..3821
                     /note="Pseudogene; similar to reverse transcriptases"
     gene            complement(6554..8653)
                     /gene="F7N22.18"
     CDS             complement(join(6554..6961,7363..7584,8127..8253,
                     8487..8653))
                     /gene="F7N22.18"
                     /codon_start=1
                     /evidence=not experimental
                     /protein_id="AAC13576.1"
                     /db_xref="GI:3047062"
                     /translation="MGLLLHSKINRLKASDNHANPISRRFNNLTLDEFLDSPGHASLT
                     RLDPNRPPGTIWFNEDSSVVGTVCSIFERDFKDPHAYWTMTPQPVITCWFETFACEKT
                     EETPDFLVLLDETHRKADRTFIDGKSQEIYKEVSSRIAEETQLCGSESSIASGG
                     LSVQAMMIYTHVAPRKKGRIYVVGSLHLEASSAHTCLPLPRDDPVLLSEKLAVAEAC
                     IANQAEKINSFGVYFDYLADKDPVFAAIFRGASSTRTKPTSANRPEVANL"
```

```
Qy  2281 ACAGCAAGCTGCAGAAGTGGTGCGCAAGAAAACAGGATAAGTTACTCGCCAAGTGCCTCAGGG 2340
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 53347 ACAGCAAGCTGCAGAAGTGGTGCGCAAGAAAACAGGATAAGTTACTCGCCAAGTGCCTCAGGG 53288

Qy  2341 CTATCAAGTTCTGAAGGACAAGATCAGCTGCTCCTTCCCACTACAACTATTCCGCAAT 2400
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 53287 CTATCAAGTTCTGAAGGACAAGATCAGCTGCTCCTTCCCACTACAACTATTCCGCAAT 53228

Qy  2401 GACAGCTCCCTCAGGACAATGCCTTCGAGGAGATATGACGCGGCCCGAGCCTAGAGAGCAGA 2460
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 53227 GACAGCTCCCTCAGGACAATGCCTTCGAGGAGATATGACGCGGCCCGAGCCTAGAGAGCAGA 53168

Qy  2461 AGATTCTGCATGTCCCTGCGGAGGCATTCATCATTCGAGCCTCGTGAATCTAGGAAGAATA 2520
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 53167 AGATTCTGCATGTCCCTGCGGAGGCATTCATCATTCGAGCCTCGTGAATCTAGGAAGAATA 53108

Qy  2521 GGAGAACGACACTCACTCAGCAGCAGGACGACTTCTCGTAGTT 2580
         |||||||||||||||||||||||||||||||||||||||||||
Db 53107 GGAGAACGACACTCACTCAGCAGCAGGACGACTTCTCGTAGTT 53048

Qy  2581 TACGGCACCGGCGGTGCTGGCCCGCAATAGAGAAGAGAGGTCGAGTATCCTCAGAGCCGGTG 2640
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 53047 TACGGCACCGGCGGTGCTGGCCCGCAATAGAGAAGAGAGGTCGAGTATCCTCAGAGCCGGTG 52988

Qy  2641 CTGGCCGCCCAGAGAGCTGATGAGATCGAGTACCCACACATGCTGGAGCTGATACGGAACATG 2700
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 52987 CTGGCCGCCCAGAGAGCTGATGAGATCGAGTACCCACACATGCTGGAGCTGATACGGAACATG 52928

Qy  2701 GCGGTTCGCTCTATGGCTTGGGAGCAATCACAGGCAGCCAGCCATTGATTGACTACCAACTTCGTTCAT 2760
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 52927 GCGGTTCGCTCTATGGCTTGGGAGCAGCCAATCACAGGCAGCCAGCCATTGATTGACTACCAACTTCGTTCAT 52868

Qy  2761 TATTCGACTGAGGTAAGGCCCTCACTTCACCATTATATCATCTCTTGTGATTTGT 2820
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 52867 TATTCGACTGAGGTAAGGCCCTCACTTCACCATTATATCATCTCTTGTGATTTGT 52808

Qy  2821 TCTTTATTTGTTTCAGTGATTGGATTGTCCTGAGTACTCTCTTCCAAGTTTATTCACA 2880
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 52807 TCTTTATTTGTTTCAGTGATTGGATTGTCCTGAGTACTCTCTTCCAAGTTTATTCACA 52748

Qy  2881 CAGTGGACTGTGTGATTTAAGTTTGGGGGAGGGCTCAGGAAGTATGTTGCATTGTATATA 2940
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 52747 CAGTGGACTGTGTGATTTAAGTTTGGGGGAGGGCTCAGGAAGTATGTTGCATTGTATATA 52688

Qy  2941 TTTTTAAGTCTGCATTCATCTAAGCATAGAAAAACCAAAAAAATTAAAAATTTCAGA 3000
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 52687 TTTTTAAGTCTGCATTCATCTAAGCATAGAAAAACCAAAAAAATTAAAAATTTCAGA 52628

Qy  3001 AAATGATTTCACAAAAAAGAGTGTTCATGTAGTTGCATTACATTTAGGATCAAGTCTAG 3060
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 52627 AAATGATTTCACAAAAAAGAGTGTTCATGTAGTTGCATTACATTTAGGATCAAGTCTAG 52568

Qy  3061 AGTGTTTCATTTAGGATTGTTGCATATGCATAGGGGATAATGATGAGATAGCCTTGTAAG 3120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 52567 AGTGTTTCATTTAGGATTGTTGCATATGCATAGGGGATAATGATGAGATAGCCTTGTAAG 52508

Qy  3121 CA 3122
         ||
Db 52507 CA 52506
```

```
misc_feature    RWATRHQLLWALALEQKPVLDPLHQQHSNP"
                11245. .14802
gene            /note="Pseudogene; similar to reverse transcriptases"
                27112. .28030
                /gene="F7N22.3"
CDS             join(27112. .27748,27822. .28030)
                /gene="F7N22.3"
                /note="contains similarity to peptidyl-prolyl cis-trans
                isomerase (Pfam: pro_isomerase.hmm, score: 23.86 and
                28.41; coded for by A. thaliana cDNA N97251; coded for by
                A. thaliana cDNA T43563"
                /codon_start=1
                /protein_id="AAC13578.1"
                /db_xref="GI:3047064"
                /translation="MASSSLIPPILTRRNLLSTTIATVTPPPAKPPSPDITITDR
                VFLDFSLCPTYFRSDPSATLSSTTPCSDSTPLGRVVLGLYGRHVPITVSTFKRMCTSS
                STSYKNTPVHKIFPGQYFLAGRQGGRRDTAEVGSYSLRDLPRNTDVVNSKAFLIPHAR
                AGVVSLCLSENDDDDIRLDPDYRNVEFLITTGPGPSPQLDGGNIVFGTVLEGLDVVT
                SISSIPTYKPSENIKQFNDFAEFLGDERAQNARSLWNRPLKTVFISGCGELKVTNPSL
                SPTLP"
gene            32752. .33612
                /gene="F7N22.4"
CDS             join(32752. .32937,33015. .33103,33358. .33424,33502. .33612)
                /gene="F7N22.4"
                /note="contains similarity to human OS-9 precurosor
                (GB:U41635)"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAC13579.1"
                /db_xref="GI:3047065"
                /translation="MVMLDKHGNRFLCYLPKEEKATSGWTSSQQNISTVMMETQOLVK
                LKTPDELLQPLSEKCLFRQEGWWSYEFCHQKYRQLHVEDENKVGFYWHSHYYTNGT
                TCDLTGSPREVEVRFVCAETRAMVTSITELSTCKYALTVQCPTLCKHP"
misc_feature    complement(35192. .39289)
                /gene="F7N22.17"
gene            complement(39286. .40766)
                /gene="F7N22.17"
CDS             complement(join(39286. .40383,40614. .40766))
                /gene="F7N22.17"
                /note="similar to Arabidopsis thaliana retrotransposon
                Tal1-1 (GB:(L47193)"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAC13575.1"
                /db_xref="GI:3047061"
                /translation="MAMLSEGDFLGDFVPVLHQLISPRWSAFPTIGIRFFLWNFDGGY
                SLPNSYNAAYVNRFSVVTTVNPRKQNRALIGQMSRVWGFPDSCVGRILERGRVQFK
                FQTEEAAMNLVLRRGPWSFNDWMLSIHRWYPNISETELKIIPFWYQITGIPLLFLTNAM
                ARCVGNRLGIVSTVDFDENSNHVGFVRVKIDWNLDDPLRFQRNFQFAVDENTVIKYRF
                ERLRNFCSKCGSLKHDVKECVLAFDDEDPAEISDDDNDDNQGDNKDKDMSDTDTLQ
                TVDPATLIPGLTQPFSTGNKGPNTQNTTVFSVPSVFEDTDLTVERLRYLHAKLARVRT
                QTPISNDLLEDSSDNAQNQFVLLKRKRVQFEDMYRKVEAADEMAVLSQLCKKERKAES
                AGSCSIHNGLDGGAGGPVPPRDPI"
misc_feature    46610. .47084
                /note="Pseudogene; similar to retrovirus-related POL
                polyproteins"
gene            57941. .59852
                /gene="F7N22.6"
CDS             join(57941. .58149,58209. .59137,59251. .59852)
                /gene="F7N22.6"
                /note="contains similarity to retrovirus-related POL
                polyproteins"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAC13580.1"
                /db_xref="GI:3047066"
                /translation="MTDINTNSKIVGNTTVVETPDEVALHLAKELEKTQITDSGMDNV
                RRNLFANDSTNTTIPSSASQGMMPDFINMERYLTEDPPIVPQGTTYVYTVGGMDTWAQ
                GDYCCKGLILNIVNDLFDLYNKASSKTLWLTLENKYKTDESGMORFSTAKFLNFKM
                VDSKPIMEQVEALQRISQEIELEGMSICNVFKTNCLIEKLPPGWLDFKNYLNFKKAM
                TFDDLVRRLMIEGNNYGAHAGAQNQGHHDVNVAEHKAKLKGKGKGFSIPQKNLKVSSTT
                NFKKSNPERKPFKGKCHHCGKIGHKGDVCKSKAKDVKSQANLTEEDMVAVVTECNMVDD
                NQVEWYYDTGCTTHICTDRTMFSTYVKNKSNEQLFMGNTAMSMIEGGLVKMSVMTVFPP
```

```
                KNVASASVVEMENPITYLVESFTTWHERLGHVNYKIMRKMQNMLLIPKFKNQEKCE
                VCVHAKLTKTPSPRVERTTEPMNLIPKFKTNQEKCELCVQAKLIKTPSPRVERTTEPL
                GLIYTDLCDLKYVQTRGGKKFFVTFIDDCTRYYVVLLHSKDEALVKFKEFTLEVENQ
                LQRTIKIVRSDRG"
gene            63170. .64252
                /gene="F7N22.7"
CDS             63170. .64252
                /gene="F7N22.7"
                /note="similar to Arabidopsis thaliana retrotransposon
                Athila (GB:X81801)"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAC13581.1"
                /db_xref="GI:3047067"
                /translation="MEALNTKVRYLEGQSASTSAPKVTGLPGKSIQNPKEYATAHAIT
                ICHDRELPTRHVPDLIIGDNDVQEGASTQIEVSVVEFNHSDGFRHLIQSTSEEKAAT
                IKRIVKRFKPTPLPSRALPWTFRKAWMERYKSAATKQLDBIEAVMPLMEVLYLIHDPH
                KDVRNLILERINMYQDLDDESDATPSRAADKRIVQKNLEDPGSFTLPCSIGELAFSDC
                LCDLGAFVSLMPLSVARRLEFIQYKPCDLTLILADRSYRKPFGMKLDLPVMINGVEVP
                TNFVVLDMEVEHKDPLILGRPFLASVGDVVDVREGKISLNLGKHNKLQFDINKIPQES
                TEDEKTSKDYRVIPGEGYETKIKKA"
gene            64919. .67482
                /gene="F7N22.8"
CDS             join(64919. .64986,66088. .66270,66336. .66473,66723. .67482)
                /gene="F7N22.8"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAC13577.1"
                /db_xref="GI:3047063"
                /translation="MIQNKETDEGPTAAAMSPTLPTRDEFERSAARRNQRRAEIARGK
                RAMSRYELIEDIKTEYEPESWRKETKLLNKSDEVTVEDPVGVTGGRAASVREVSSD
                KEETIEFLSTLQVEMYEGLTDFELDTMGESTGTVSNTDMKMIDSALIGILRLTKGKNV
                LRGDLNDSPPVMPLLIHLCGYMKWALINGKKVRGALCVGGVVTPILKVCGVPLKEVG
                LAPRRMDLDHLRRCEFSEFDMVGDFHRYRFEHSSIRIANLFPCIYATRILEGRNIDF
                KPALEDLYFEGSPPTBEISHTEGATIEDVDETYDIDEAEFDTSMYHFSEHIPPARKSK
                SLSEAHRNNSKLQKWCKKQDKLLAKCLRAIKFLKDKISCSSSTTTIPQ"
misc_feature    72089. .73326
                /note="Pseudogene; similar to retrovirus-related POL
                polyproteins"
gene            complement(84219. .86759)
                /gene="F7N22.15"
CDS             complement(join(84219. .84649,84749. .85148,85424. .85967,
                86365. .86759))
                /gene="F7N22.15"
                /note="contains similarity to retrotransposon-like


Query Match                              Score 3122;    DB 8;   Length 110157;
Best Local Similarity    100.0%;     100.0%;         Pred. No. 0;
Matches 3122;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1 TTCATATATTCGACCTCTTCTCTCTTCTTGCATCCAAACAGACAACAAGCCGGCATC   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 65081 TTCATATATTCGACCTCTTCTCTCTTCTTGCATCCAAACAGACAACAAGCCGGCATC 65140

Qy    61 GCTTTCCCTCACAACTCTCACTCGACCACCGGCTCCCTCTCCTCACTTACTGGCTTCATCG   120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 65141 GCTTTCCCTCACAACTCTCACTCGACCACCGGCTCCCTCTCCTCACTTACTGGCTTCATCG 65200

Qy   121 CTCTCATCGGCCATCTCTCAACATACTCGACCTCGCGATATCACTCGAGCTCGCCGCTTCT   180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 65201 CTCTCATCGGCCATCTCTCAACATACTCGACCTCGCGATATCACTCGAGCTCGCCGCTTCT 65260

Qy   181 CACCGGCTTCTCACCGGCCTGCTCCCGTCCTCTCCAAGGAAACAACTCGAGCTCC   240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 65261 CACCGGCTTCTCACCGGCCTGCTCCCGTCCTCTCCAAGGAAACAACTCGAGCTCC 65320

Qy   241 ATTTCACTCACTCGACCTCTCACCACCAAGCCGGCTTCCACCACTTCAGCTCTTAACCACT   300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 65321 ATTTCACTCACTCGACCTCTCACCACCAAGCCGGCTTCCACCACTTCAGCTCTTAACCACT 65380

Qy   301 CGACCACCTTCACCATCAACCAATCAACCAATCAAATCGTTTTCCTCCATTAAAGCTTGACATACT   360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 65381 CGACCACCTTCACCATCAACCAATCAACCAATCAAATCGTTTTCCTCCATTAAAGCTTGACATACT 65440
```

```
Qy   1441  AGATCAAGAAATTGGAAGAACTGTTTGGTTTCCCTAGTGGAAAGGGAACCAACCCCAGGT  1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66521  AGATCAAGAAATTGGAAGAACTGTTTGGTTTCCCTAGTGGAAAGGGAACCAACCCCAGGT  66580

Qy   1501  TTGACAGGGAAGAGCTTAAGGATTGTGGGCTACTATTGGGCTACTATTCCGGCTAAACT   1560
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66581  TTGACAGGGAAGAGCTTAAGGATTGTGGGCTACTATTGGGCTACTATTCCGGCTAAACT   66640

Qy   1561  CGACGCGGTCCAAGAGCAACCAAATCCGGAGTCGTGATTCGTGATTCGCTACTTTCAGCGGCTCGG  1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66641  CGACGCGGTCCAAGAGCAACCAAATCCGGAGTCGTGATTCGTGATTCGCTACTTTCAGCGGCTCGG  66700

Qy   1621  TTGCCAATGTTTTTACTCCAGGAGTCTACAGGCACCGTGTCTCAACACAGACATGAGAGA  1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66701  TTGCCAATGTTTTTACTCCAGGAGTCTACAGGCACCGTGTCTCAACACAGACATGAGAGA  66760

Qy   1681  TGATAGATTCAGCGCTTATAGGGATTCTCCGGCTTCCCGGCTTACAAAAGGAAAGAATGTCCTGAGAG  1740
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66761  TGATAGATTCAGCGCTTATAGGGATTCTCCGGCTTCCCGGCTTACAAAAGGAAAGAATGTCCTGAGAG  66820

Qy   1741  GAGATCTTAACGACTCACCAGTAATGCCTCTGTTGATCCATCTGTGTGGGTACATGA    1800
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66821  GAGATCTTAACGACTCACCAGTAATGCCTCTGTTGATCCATCTGTGTGGGTACATGA    66880

Qy   1801  AGTGGGCGCTGACAAACGGCAAGAAGGTAAGACGCACTATGCGTGGGTGGCGTTG     1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66881  AGTGGGCGCTGACAAACGGCAAGAAGGTAAGACGCACTATGCGTGGGTGGCGTTG     66940

Qy   1861  TGACGCCAATTCGAAAGTTGTGGAGTTCGCTCAAGGAGTAGGGTTAGCACCGGAGAA   1920
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66941  TGACGCCAATTCGAAAGTTGTGGAGTTCGCTCAAGGAGTAGGGTTAGCACCGGAGAA   67000

Qy   1921  TGATGGACTTGGATCACTTGCGCCGATGTGAGTTCTCTGAGTTTGACATGGTTGGCGACT  1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  67001  TGATGGACTTGGATCACTTGCGCCGATGTGAGTTCTCTGAGTTTGACATGGTTGGCGACT  67060

Qy   1981  TTCACCGCTACAGGTTCGAGCATTCATCGATTAGAATCGCCAACATTCTTTTCCCCTGCA  2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  67061  TTCACCGCTACAGGTTCGAGCATTCATCGATTAGAATCGCCAACATTCTTTTCCCCTGCA  67120

Qy   2041  TTTACGCTACTAGGATTCTCGAGGCAGGAACATTGACTTCAAGCCTTGAAGATC      2100
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  67121  TTTACGCTACTAGGATTCTCGAGGCAGGAACATTGACTTCAAGCCTTGAAGATC      67180

Qy   2101  TTTATTTCGAGGGCAGTCGCCAACTGAGGAGATTAGTCACACCGAGGAGTACAATAG     2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  67181  TTTATTTCGAGGGCAGTCGCCAACTGAGGAGATTAGTCACACCGAGGAGTACAATAG     67240

Qy   2161  AAGATGTTGATGAGACATATGATAGATGAGGCGGAGTTTGAGCGAAAAGCAAGACAGGACA  2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  67241  AAGATGTTGATGAGACATATGATAGATGAGGCGGAGTTTGAGCGAAAAGCAAGACAGGACA  67300

Qy   2221  TCAGTGAGCGAAGCTGCATATACCTTCCAGCGAGGAGAATCATTCGAGACGCAGGACGAACA  2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  67301  TCAGTGAGCGAAGCTGCATATACCTTCCAGCGAGGAGAATCATTCGAGACGCAGGACGAACA  67360

Qy   2281  ACAGCAAGCTGCAGAAGTGGTGCAAGAACAACAGGATAAGTTACTCGCCAAGTGCCTCAGGG  2340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  67361  ACAGCAAGCTGCAGAAGTGGTGCAAGAACAACAGGATAAGTTACTCGCCAAGTGCCTCAGGG  67420

Qy   2341  CTATCAAGTTTCTGAAGGACAAGATCAGCTGCTCCTCCTCTTCCACTACAACTATTCCGCAAT  2400
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  67421  CTATCAAGTTTCTGAAGGACAAGATCAGCTGCTCCTCCTCTTCCACTACAACTATTCCGCAAT  67480

Qy   2401  GACAGCTCCCTCCAGGACATGCCTTCGAGAGGAGATATGAGACGCGCCCGAGAGCAGA   2460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  67481  GACAGCTCCCTCCAGGACATGCCTTCGAGAGGAGATATGAGACGCGCCCGAGAGCAGA   67540

Qy   2461  AGATTCTGCATGTCCCTGCGAGGCATTCATTCGAGCCTCGTGAATCTGAATTCTAGGAGAATA  2520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  67541  AGATTCTGCATGTCCCTGCGAGGCATTCATTCGAGCCTCGTGAATCTGAATTCTAGGAGAATA  67600

Qy   2521  GGAGAACGACACTCACTCGATCTAGCAGGAGCAGGAGCAGACTTCTGCAGTCGTGAGTT      2580
```

```
Qy    361  CGACCGCTGAACACTTATCACCTTCAAGCTCCTCATCTCTTCATCGTTCCAACACCGCT   420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 65541  CGACCGCTGAACACTTATCACCTTCAAGCTCCTCATCTCTTCATCGTTCCAACACCGCT  65500

Qy    421  GCTCTCATCCCCACGAAAGCTTGTCATCATCGAACCTCACTCTTCACTCTTCACTTCGATTC  480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 65501  GCTCTCATCCCCACGAAAGCTTGTCATCATCGAACCTCACTCTTCACTCTTCACTTCGATTC  65560

Qy    481  AGCAACCAAACTCGACCTCGTCTTCTTTGCCACTCATAGTCACTCCTCCTCCACCA       540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 65561  AGCAACCAAACTCGACCTCGTCTTCTTTGCCACTCATAGTCACTCCTCCTCCACCA       65620

Qy    541  TCTTCATCATCTCCCTTACTCGACCACCGTGCGTCTCGCTCCACCCATTGCCATTAAAAG   600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 65621  TCTTCATCATCTCCCTTACTCGACCACCGTGCGTCTCGCTCCACCCATTGCCATTAAAAG   65680

Qy    601  CTCACTCGATTGTCAAAGAGAGTCAAGCTCAACCACCACCGCGACCGCGTTT          660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 65681  CTCACTCGATTGTCAAAGAGAGTCAAGCTCAACCACCACCGCGACCGCGTTT          65740

Qy    661  CCCTCTACACATTCAACACTCGACCACGGTGCTACCACCGTGCTACCACCGTGTTCAC     720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 65741  CCCTCTACACATTCAACACTCGACCACGGTGCTACCACCGTGCTACCACCGTGTTCAC     65800

Qy    721  CATACACTCGACCAACAACTCTCAAAGTAAAAAAAAGAAAAAAAGAAAAAGTCAAAACCGA  780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 65801  CATACACTCGACCAACAACTCTCAAAGTAAAAAAAAGAAAAAAAGAAAAAGTCAAAACCGA  65860

Qy    781  CAGTTCACTCACTCAACCGGTTTACTCGACCGGTTAGTTTTGGTGTTTTTAGATTGTGTTTTTTGGTTTT  840
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 65861  CAGTTCACTCACTCAACCGGTTTACTCGACCGGTTAGTTTTGGTGTTTTTAGATTGTGTTTTTTGGTTTT  65920

Qy    841  GCTATTACTAACATATTATCTTTGAGTTTCGTCGTTTTTTAGGTTTCATCATG       900
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 65921  GCTATTACTAACATATTATCTTTGAGTTTCGTCGTTTTTTAGGTTTCATCATG       65980

Qy    901  AGTAACTACAGTGGAAAATCCTCTATGGACCCTGATTATTAATGTCGATGAAGCTAAGTCC  960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 65981  AGTAACTACAGTGGAAAATCCTCTATGGACCCTGATTATTAATGTCGATGAAGCTAAGTCC  66040

Qy    961  TGGTCCACTAGACCGGAGTGAGTGAGAGCAACATGTTTACGAGAGCTATAGGGATGAATTTGAA  1020
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 66041  TGGTCCACTAGACCGGAGTGAGTGAGAGCAACATGTTTACGAGAGCTATAGGGATGAATTTGAA  66100

Qy   1021  CGCTCTGCAGCTCGACGTAATCAAAGAAGAGCTCGAAATCGCTAGACGACGAAAGAGGGCGATG  1080
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 66101  CGCTCTGCAGCTCGACGTAATCAAAGAAGAGCTCGAAATCGCTAGACGACGAAAGAGGGCGATG  66160

Qy   1081  TCGAGTAGATATGAGCTGATTGAGCTCAAAACTGAGTATGAGCCAGAGTCATGG       1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 66161  TCGAGTAGATATGAGCTGATTGAGCTCAAAACTGAGTATGAGCCAGAGTCATGG       66220

Qy   1141  CGCAAGGAGACGAAGCTACTGAACAAATCCGACGAGGTTACAGTGGAGGTATATCAGA    1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 66221  CGCAAGGAGACGAAGCTACTGAACAAATCCGACGAGGTTACAGTGGAGGTATATCAGA    66280

Qy   1201  TTCTTTGAGATGAATGAATCTCGGGGAACGAGCTATCCCTGATATGAGACTTTAGCCCAG   1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 66281  TTCTTTGAGATGAATGAATCTCGGGGAACGAGCTATCCCTGATATGAGACTTTAGCCCAG   66340

Qy   1261  TTTGGGGTTACTCGAGGACGTGCAGCATCTGTTCGAGAAGTGTCATCTGATAAGGAGGAGA  1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 66341  TTTGGGGTTACTCGAGGACGTGCAGCATCTGTTCGAGAAGTGTCATCTGATAAGGAGGAGA  66400

Qy   1321  CAATCAGTTGTTCTTTCCACACTGCAAGTGGAAATGTATGAGGGACTTCACAGACTTTGAGC  1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 66401  CAATCAGTTGTTCTTTCCACACTGCAAGTGGAAATGTATGAGGGACTTCACAGACTTTGAGC  66460

Qy   1381  TGGATACCAATGGGGTTAGGGCTTCTTGACGTTCTTGACGTTCTTGACTTAGTGATGAACAGCGGTACCAGAGATTT  1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 66461  TGGATACCAATGGGGTTAGGGCTTCTTGACGTTCTTGACGTTCTTGACTTAGTGATGAACAGCGGTACCAGAGATTT  66520
```

```
Db  67601 GGAGAACGACACTCACTCGATCTAGCAGCAGAGGACGACGACTTCTGCAGTCTCGTAGTT 67660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2581 TACGCGACCGCGTGCTGGCCCGCAATAGAATAGAAGAAGAGGTCGAGTATCCTCAGAGCGGTG 2640
Db  67661 TACGCGACCGCGTGCTGGCCCGCAATAGAATAGAAGAAGAGGTCGAGTATCCTCAGAGCGGTG 67720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2641 CTGGCCGCCCACAGAGCTGATGAGATCGAGTCGAGTCTGGAGCTGGATGAGCTGGATGACCATG 2700
Db  67721 CTGGCCGCCCACAGAGCTGATGAGATCGAGTCGAGTCTGGAGCTGGATGAGCTGGATGACCATG 67780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2701 GCGGTTCGTCTCATGGCTTGGGAGCAATCACAGCGAGCCATTGACTACCAACTTCGTTCAT 2760
Db  67781 GCGGTTCGTCTCATGGCTTGGGAGCAATCACAGCGAGCCATTGACTACCAACTTCGTTCAT 67840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2761 TATTCGACTGAGGTAAGCGCCTCACTTCACCATTATTATTATCATCTCTTGTGATTTGT 2820
Db  67841 TATTCGACTGAGGTAAGCGCCTCACTTCACCATTATTATTATCATCTCTTGTGATTTGT 67900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2821 TCTTTATTTTGTTTCAGTGATTGATTGATTGTTCCTGAGTACTCTTCTTCCAAGTTTATTCACA 2880
Db  67901 TCTTTATTTTGTTTCAGTGATTGATTGATTGTTCCTGAGTACTCTTCTTCCAAGTTTATTCACA 67960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2881 CAGTGGACTGTGTGATTTAAGTTTGGGGGAGGGCTCAGGAAGTATGTTGCATTGTATATA 2940
Db  67961 CAGTGGACTGTGTGATTTAAGTTTGGGGGAGGGCTCAGGAAGTATGTTGCATTGTATATA 68020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2941 TTTTTAAGTCTGCATTCATCTAAGGCATAGAAAAACCAAAAAAAAATTAAAAATTTCAGA 3000
Db  68021 TTTTTAAGTCTGCATTCATCTAAGGCATAGAAAAACCAAAAAAAAATTAAAAATTTCAGA 68080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3001 AAATGATTTCACAAAAAAAGAGTGTTCATGTTCAGTTGCATTACATTTAGGATCAAGTCTAG 3060
Db  68081 AAATGATTTCACAAAAAAAGAGTGTTCATGTTCAGTTGCATTACATTTAGGATCAAGTCTAG 68140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3061 AGTGTTTCATTTAGGATTGTTGTTGCATATGCATAGGGGATAATGATGAGATAGCCCTTGTAAG 3120
Db  68141 AGTGTTTCATTTAGGATTGTTGTTGCATATGCATAGGGGATAATGATGAGATAGCCCTTGTAAG 68200
          ||
Qy   3121 CA 3122
Db  68201 CA 68202
```

e-mail: rwilson@watson.wustl.edu

MAPPING: Clones were assigned to the YAC map by hybridization by
M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted
by M. Marra, WashU, to pick the best candidates for sequencing.

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' clone is T5H22 ;3' cosmid is T25H8. Actual start of this
clone is unknown ; actual end is at 74475 of T7M24.

NOTES:

Coding sequences below are predicted from computer analysis, using
the program Genefinder(P. Green and L. Hillier, ms in preparation).

```
FEATURES          Location/Qualifiers
     source       1..74475
                  /organism="Arabidopsis thaliana"
                  /mol_type="genomic DNA"
                  /cultivar="Columbia"
                  /db_xref="taxon:3702"
                  /chromosome="IV"
                  /map="unknown"
                  /clone="T7M24"
     gene         4936..7830
                  /gene="T7M24.7"
     CDS          4936..7830
                  /gene="T7M24.7"
                  /note="contains similarity to reverse transcriptase (Pfam:
                  rvt.hmm, score 19.29)"
                  /codon_start=1
                  /evidence=not_experimental
                  /protein_id="AAC26250.1"
                  /db_xref="GI:3319352"
                  /translation="MVLELKNCYVVPAINKNIISVSCLDMEGFHFSIKNKCCSFDRDD
                  MFYGSAPLDNGLHVLNQSMPIYNIRTKKFKSNDLNPTFLWHCRLGHINEKHIQKLHSD
                  GLLNSFDYESYETCESCLLGKMTKAPFTGHSERASDLLGLIHTDVCGPMSTSARGNYQ
                  YFITFTDDFSRYGVYVLMCHKSKSFENFKFQNEVQNQPGKSIKALRSDRGGEYLSQV
                  FSDHLREGIVSQLTPPGTPQWNGVSERRNRTLLDMVRSMMSHTDLPSPFWGYALETS
                  AFMLNRCPSKSVEKTPYEIWTGKVPNLSFLKIWGCESYARLITDKLGPKSDKCYVG
                  YPKETKGYYFHPTDNKVFVVRNGAFLEREFLSKGTSGGKVLLEEVREPGDVPTSQE
                  EHQLDLRRVVEPLLVEPEVRRSERSRHEPDRFRDWVMDDRALFMESDEPTSYEEALM
                  GPDSDKWLEAAKSEMESMSQNKVWTLVDLPDGVKPIECKWIFKKKIDMDGNIQIYKAG
                  LVAKGYKQVHGIDYDETYSPVAMLKSIRILLATAAHYDYEIWQMDVKTAFLNGNLEEH
                  VYMTOPEGPTVPEAARKVCKLHRSIYGLKQASRSWNLRFNEAIKEFDFIRNEEEPCVY
                  KKTSGSAVAFLVLVDDILLLGNDIPLLQSVKTWLGSCFSMKDMGEAAYILGIRIYRD
                  RLNKIIGLSQDTYIDKVLHRFNMHDSKKGFIPMSHGILSKTQCPSTHDERERMSKIP
                  YASAIGSIMYAMLYTRPDVACALSMTSRYQSDPGESHWIVVRNIFKYLRRTKDKFLVY
                  GGSELVVSGYTDASFQTKDDFRSQSGFFCLNGGAVSWKSTKQSTVADSTEAEYI
                  AASEAAKEVVWIRKFITELGVVPSISGPIDLYCDNNGAIAQAKEPKSHQKSKHIQRRY
                  HLIREIDRGDVKISRVSTDANVADHFTKPLPQPKHESHTTAIGIRFIKM"
     misc_feature 15079..19726
                  /note="pseudogene"
     misc_feature complement(21094..21816)
                  /note="pseudogene"
     gene         41914..45895
                  /gene="T7M24.5"
     CDS          join(41914..42178,42523..42709,44317..44441,44538..45141,
```

```
RESULT 4
T7M24
LOCUS       T7M24                 74475 bp    DNA     linear   PLN 15-JUL-1998
DEFINITION  Arabidopsis thaliana BAC T7M24.
ACCESSION   AF077408
VERSION     AF077408.1  GI:3319359
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 74475)
  AUTHORS   Harmon,G., Langston,Y., Stoneking,T., Drone,K. and Ames,M.
  TITLE     The sequence of Arabidopsis thaliana T7M24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 74475)
  AUTHORS   Washington University Genome Sequencing Center.
  TITLE     The Arabidopsis thaliana Genome Sequencing Project
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 74475)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     Submitted by:
            Genome Sequencing Center
            Department of Genetics, Washington University,
            St. Louis, MO 63108, USA
```

YETLAQLGLLEDSLTDFELDTMGLGFLTFSVDEQRYQLSIKKLEELFGPPSGNGTKPK
FERKELKDLWATSGNNVPLNSARSKNOIRSPVIRYFQRSIANVLYSRESTGVFNTD
MEMIDSALTGILRRTKGKNVLRGNLNDAPPVMPLLIHLCGYRKWALTNGKKKGSPPTE
EISHTEGAIKEDVDVTVEIDEVFEDTSMYHFSEHIPPAREIKSLSEAQKNNSKLQKWC
KKLDKLLAKCLRAIKFLKDKLSCSSSTTAIPQQLPQDMPSRRYDAPKPSRRPEPSE
QEIPHVPARHSSEPRESGRKRRTTLTQSSSTSKRLLQSRSLRDRGAVSRRREVEYP
HSGAGRHRADEIEYPPAGADTEHGGSSMACEQ"

gene   66446..73816
              /gene="T7M24.1"
CDS   join(66446..66721,66811..66848,66888..67029,67190..67535,
              72840..73180,73358..73816)
              /gene="T7M24.1"
              /note="contains similarity to maize transposon MuDR
              mudrA-like proteins"
              /codon_start=1
              /evidence=not experimental
              /protein_id="AAC26252.1"
              /db_xref="GI:3319364"
              /translation="MKGASIEEGDANQNMDEVDVNAEEDTNEKKDEDEMSRFDCFDD
TDGASDGDEESSVCTTYLKIKTSVDLLIMEMVTICLNYGKVHQDDNKFIKPDFNVGRST
LELVRATPTGVGSCFEIRVYISDHTCSVTARSNRARQATHAILGQIYKESVNYWMSHR
TLKFARHLVRGRDESEYDQLPSYLYMIRKALHLSRLEVDLDRFKYVFIAFGASIDGF
PFIRKVIVDEIFLKGKYLGTLLLATTQDENFNIFPLAFAIVDTKNDASWKAIGSVYP
LSSRGSCTYHLHKNILLRFKGSESFGLVKKAATAFRLSDFNVLFQQIHELNLEHSYL
ERADLSMLNRVHFSGDRYNFTTSNNVETINKVLRPYRNVHIVALLDKSRVLRASILKV
QEIDPHQYEVRSSSTINVVNLSQNKCTCRMFDLEKLPCIHALASAVEMAKVSRITKCHP
YYRKDYLCNRYDKSVMSQESCTLPDIVLQKICHSPDVRPQPGRPKTTRVKSALEVALD

                                                                    Query Match   65.0%; Score 2029; DB 8; Length 74475;
                                                                    Best Local Similarity 81.3%; Pred. No. 0;
                                                                    Matches 2507; Conservative 0; Mismatches 495; Indels 81; Gaps 10;

45538..45895)   Qy   102 CACTTACTGGCTTCATCGCTTCTCATCGCCATCTCAACATACTCGACCTCGCGATATC 161
  /gene="T7M24.5" |||||| |||| ||| |||| || || |||||| ||||||| | || |||||| |
  /note="contains similarity to Arabidopsis thaliana   Db   60404 CAAACACTCGATCTCATCGACCACTCTTCGCCACCACCACCTCTTCAACATACTCGACCTTTC 60463
retrotransposon Athila (GB:X81801)"
  /codon_start=1   Qy   162 ACTGGAGCTCGCCGCTTCTCACCGCCTCATCGTTCACCGCCTCGCTCTCCTCTCCAAG 221
  /evidence=not experimental ||||||||| |||||| | ||||||||||| ||||||||| ||||| ||||
  /protein_id="AAC26249.1"   Db   60464 ACTCGACCACAC---TCTCTTCGCCCTCCCATCATCACGGCCGTCGCTCCTCTCCAAG 60520
  /db_xref="GI:3319361"
  /translation="MADEQNOQNGPANIGAGDAPRDHRQRKRIAPPAIQNNNFEIKSG   Qy   222 GAAACAACTCGAGCTCTCCATTCACTCACTCGACCTTCTCACCACCAGCCGGCTTCACCA 281
LISMIQGNKFHGLPMEDPLDHFDEFDRLCNLTKINGVSEDGFKLQGWELVENLAQSDG ||||||||||||||| ||||| ||||||| ||||||| ||||| |||||||||
NYNEDCDRTVRSTADSDDKHRKEIKALNDKLDMILFSQQKHLHFLVDDEVFEDFPEKD   Db   60521 GAAACAACTCGACCTCACCATTTCATTCACTCGACCTTCTCACCACCAACCAATCAAATCAACCA 60580
EQWRKSFFFRVNELTFGEKTCLFVSEWSARAASVDLLPISKTFASHFSSFVRQDLGW
ESLSSDDVRASGKRLRSRTDLAISSPPFERIFSEEBMSAPSYRAEKKMERRARKQEK   Qy   282 CTTCTAGTCTTAACACTTGACCACCTTCACCATCAACCAATCAAATCGTTTTTCTCCTC 341
DRLKSDKVVMTGTLEDSKLQAANKAKQPRAEBDGGARQKPSTLASKNDAPAPDSRTPS ||||||| ||||||||||||||||||||||| ||| |||||||||||||||||||
KATSSRDPSKRYGDKKRKKQSDRDARSPARSVRRTEERDASPSQQKEKGKKGDSQDLV   Db   60581 CTTCTCGCTCTTAACCACTGACCATCTCACCATCAACCAATCAAATCGTTTTTCTCCTC 60640
VLEIESFKQNEEENSRMVDRANSVLSMREAZARVLAIDIANTDLAAKLESGKNAYLA
AIDNENRARAELLACEBKLRKLEEGQAALLADARREERRKVRAQFKDFTSKYQSFYAE   Qy   342 CATT------AAAGCTTGACATATCGACCGCTTCACTCGACCTTATCACCTTCAAGCTCCT 393
SEEV" |||| ||||||| | || ||||||||||||| | |||||||||||
                                                                    Db   60641 CATTCAAGTTCCAAAACTGGACCGCGCCTTCACTCGCTCTCATCCCCACGAAAGCTTCA 60700
misc_feature complement(46760..47857)
                /note="pseudogene"   Qy   394 CATCTTCTTCATCGTTTCCAACACGACCGTCATCACGAACTCACTCGTGTCATCACCTT 453
gene   50913..54910 |||||||||||||||| | |||| || ||| || |||||| ||||||||||||
                /gene="T7M24.4"   Db   60701 CCGTCTACACACTCACTCGCCCTTTGCCCGCCGGTCTCCTTCTTCACCATCACACGAACTCACCT 60760
CDS   join(50913..51681,51997..52077,52491..52774,52906..53838,
                53999..54163,54221..54910)   Qy   454 CTCACTCATCACCAGTTCACTCGATTCAGCAACCAAACTCGACCTCGTCTCTCTTGCCAC 513
                /gene="T7M24.4" ||||||||||||| | || |||||| |||||||| |||||||||| ||| ||||
                /note="contains similarity to reverse transcriptase (Pfam: Db   60761 CTCACTCCTCATCTTCACTCGACCGGTGTAAGAGAAGCTAACCAAACAATCCGATGAA 60820
                rvt.hmm, score: 33.26)"
                /codon_start=1   Qy   514 TCATAGTCACTCGATCTCTCCTCACACATTCCTTCATCATCCTTACTCGACCCCCTGCGG 573
                /evidence=not experimental |||||| ||||| || || ||| | |||||| | | ||||||||||||||
                /protein_id="AAC26251.1"   Db   60821 CCGTCTACACACTCACTCGCCCTTTGCCCGCCGATCAAGTACTTGAAGGTTCAGC----- 60875
                /db_xref="GI:3319363"
                /translation="MGKSSKTGKSKKDKGIVEETLVEETMINEKTRVGTMLDGTDSV   Qy   574 TCTCGGTCCACCATTGCCATTAAAAGCTCACTCGATTGTCAAAGAGAAGAAGAGTGAAG 633
GPETGKKSINEBARESSGAARDNTDNRANNTPTAQEQWDLMKGMWAQLATLTKAIVANPV ||| ||||||||| ||||| |||| ||||| || || | |||||||| |||
VQRVDNQKDDDELEVEVNFPQPNHGRKRGDYLSLLEHVSRLGTRHFMGSTDPIVADEWR   Db   60876 -CTTGCTCATTGGGTTCGTTTGGAGAATCACTCTGT-----CACAACATATCCTCAAC 60929
SRLKRNFKSTRCPEDYQRDIAVHFLERDAHNWWLTVEKRGDEVRSFADFEDEFNKKY
FPPEAWDRLECAYLDLVQGNRRTVREYDEEFNRLRRYVGRACGRCGSKDHAIQICPRME   Qy   634 CTCAACCACCGCCACTCAACTCGACGTTCCCTACCACAATCAACCATTCAACCACCAGCCGTCT 693
LGQSKVLGPTDVRKTAFLTRYGHFEFVVMPFGLTNAPTAFMKMMNGVFRDFLDEFVI ||| || || |||| | |||||||||| |||| | ||||| ||| | || ||
FIDDILVYSKSWEAHQEHLRAVLEQLREHELFAKLSKCSFWQRSCEKNFLELKAMLTN   Db   60930 CAAAGAGTTCACTCGACACCACCGCGAATCTTCTTCGCTATCACCGGTTCACAA--GCTCA 60987
APVILVLPEGEPYTVYTDASVVGLECVLMQKGSVIAYASRQLRKHEKNYPTHDLEMAA
VVFALKIWRSYLYGAKVQIYTDHKSLKYIFTQPBLNLRQRRWMELVADYDLDIAYHAG   Qy   694 ACCATTCTCCACCACCGGCTCTTGTTCACCATACACTCGACCACCACACACTCTCAAAGTAAAA 753
KANOVADALSRRSEVEAERSQVDLVNMWGTLHVNALSKEVEPLGLGAADQANLLSRI
RLAQERDEIKGWALNNKTEYGINNGTIVVNGRVCVPNNRALKEEILREAHQSKFSI
HPGGNKIYRDLKRYHVWGMKKDVARWVAKCPTCQLVKAEHQVPSGLLQNLPIPEWKW
DHITMDFFWKAFQKALGTRVNLSTAYHFPQTDGQSERTIQTLEDMLRACALDWGGNWEK
YLRLALYGRACRTPLCWTPVGERRLFGPIIVDETTERMKFLKIKLKEAQDRQKSYANK
RRKELEFQVEDLVVLKAMTYKGAGRFTSRKKLSPRYVGPYKVIERVGAVAYKLDLPPK
LNAFHNVFHVSQLRKCLSNQEESVEDVPPGLKENMTVEAWPVQIMDRMTKGTRGKSRD
LLKVLWNCGGREQYTWETENKMKANFSEWFKEMGKGQLDADSRTNPIQGGETCNARDP
Q"

gene   58231..63074
                /gene="T7M24.3"
CDS   join(58231..58457,58540..58666,58733..59640,60165..60567,
                60875..61076,61183..61571,61686..62153,62436..63074)
                /gene="T7M24.3"
                /note="contains similarity to Vicia faba
retrotransposon-like gene (GB:AB007467)"
                /codon_start=1
                /evidence=not experimental
                /protein_id="AAC26248.1"
                /db_xref="GI:3319360"
                /translation="MKQMVQQLLQCQASSSTEIAKKLPELHHKLDCSFNDLNAKVEAL
NTKVRYLEGQSASTSAPKVTGLPGKSIQNPKDDVQBGDASTQIEVSVVERNHYAGSRH
LIQSTSEEKAAIIEKMRERYKSVAAKQLDEIEAVMPLIEVNLLIHDPHKVVRNLILER
IKMYHDSNDDSDATPSRAADKRIVQEKFGDPGSFTLPCSIGELAFSDCLCDLGASVSL
MPLSMARRLEFIQYRPCDLTILILADRASRKPFGMLKDLPVMINGVEVPIDFVVLDMEV
EHKDPLILGRPFLASVGAVIDVREGKISLNLGKHIMLQFDINKTPESKEDEKTSGDD
RVIPGEKYETEKVKELKKRSDKQEETIERLAHSVEBLRSKLNQLQKEAQPKGEIDTIL
RKKFTSRWFBEIDYPPEEKEAYFEERGIEQEDIVGPAADAHVFHARDQETRNLFTPPL
HHSTSSPYYPHHSTSPSP5IINTRLLPFTFIPTTLDSHFSHRPSLSSSPNTRSRP
LFATTTSHSSQPHSTTTLFASPSSRPAPSLQGNNSTSPFHSLDLLYHQPCSFGFVWRIT
IMSNYNGESSMDPDYNVDEAESWSTRPKREHVYESFKGELERSAARNQRRAKIARG
KRAMSSRYELIDEDIETEYEPESWRMETKLLNKPDEVTVEBYIRFEMNDFWGTRYPC

```
Db  60988  CTCGACCGCTTCGCTCAAAGTGTACTCGACCCTTCACCATCAAGCAACAAAAACTCGA  61047
                   ||||||  ||  ||||       |||    |||||||||||||||||||||| ||||
Qy    754  AAAAAAGAAAAAAAAGTCAAACACCGACAGTTTCACTCAACCGGTTTACTCGACCGGTAC  813
           |||||||||  ||  ||||||||   |||||||||   ||||| || ||||||||||||
Db  61048  TTTCCAAAACCACAAAATCACTCGGCCGCGGTTT-ACTCAATAACATTACTTGACCAATTT  61106

Qy    814  GCTGGTTTAGATTGTGTTTTTGCTATTACTAACATATTAACGTTTATCTTTGAG  873
           |||||  |||||| ||||||||||   |||||||||| ||||| ||||||||||
Db  61107  ACTCTGTAGTTTATGTTTTTGGTTTTGTTGGGACTAACGTATTAAACGTTTATCTTTGAG  61166

Qy    874  TTTCGTCTGTTTTTAGGTTCATCATGAGTAACTACAGTGGAAAATCCTCTATGGACCCT  933
           ||||| ||||||||||||||||||||||||||||||  |  ||||||  ||||||||||
Db  61167  TTTCATCTATTTGATCATTTCAGTTTCATTGGGACTAACGTATTAAACGGAGAATCCTCTATGGATCCT  61226

Qy    934  GATTATAATGTGGATGAAGCTAAGTCCTGGTCCACTAGACCGGAGTTGAGAGCAACATGTT  993
           |||||||| || ||||||| ||||||| ||||||||||  |||||||||||||||||||
Db  61227  GATTATAATGTAGATGAAGCTCAGTTCGTCGGTCCACTAGACCCAAGAGACGAACACACGTT  61286

Qy    994  TACGCTACTAGGAGGATGAATTGAACGCTCTCGCAGCTCGACGTAGATATGAGCTGAATCAAAGAGAGACT  1053
           |||||||||||||||  ||||||||| |||||||| ||||||||||||||||||||||||
Db  61287  TACGAA-TTCAAAGCTTCAAAGAGAGAAGAAGCTCTCGCAGCTCGACGTAGATATGAGCTTGAAGATGGA  61346

Qy   1054  GAAATCCTAGAGGAAAAGAGGGCGATGTCAGGTAGATATGAGCTGATTGATGAGGATATC  1113
           ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db  61347  AAAATCGCCAGGAAGAGAGAAGAGAGCGATGTCAAGTCAGTCACAGAGAAGAGAAGAGAGCT  61406

Qy   1114  AAAACTGAGTATGAGCCGAAGGAGCCAGTCAACGAGCGAAGCTACTGACAAATCCGAC  1173
           |||||||||||||||||||||| ||||||||||||||| ||||||||||||||||||
Db  61407  GAAGTTGAGTATGAGCCGAAGGAGCTGCTGAACAAACCCGAC  61466

Qy   1174  GAGGTTACAGTGGAGGAGTAATGATCAGATTCTTTGAGATGAATGAATGCTCTGGGGAACGAGG  1233
           ||| |||||| |||||||||||||||||||||| |  ||||||||||||||||||||||||
Db  61467  GAGGTTACGGTGGAGGAGTAATGATCAGATTCTTTGAGATGAATGAATGCTCTGGGGGACGAGA  61526

Qy   1234  TATCCCTGATATGAGACTTTAGCCCAGTTGGGGTTACTGGAGGACGTGCGCATCTGTTC  1293
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61527  TATCCCTGCATGAGACTTTAGCCCAGTTGGGGTTACTGGAGGACGTGCAGCATCAGTTC  61586

Qy   1294  GAGAAGTGTCATCTG---------ATAAGGAGGAGACA  1322
                  || |||||| |||||||||||||||||
Db  61587  GAGAAGTGTCATCTGGAGACACTCATGTCTTACCCTTACCGGAGTCTTACCGTATAAAGAGGAGAGG  61646

Qy   1323  ATCGAGTTTCTTTCCACACTGCAAGTGCAAGTGGAAATGAATGTGAGGACTTTGAGCTG  1382
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db  61647  ATCGAGTTTCTTTCCACACTGCAAGTGCAAGTGGAGATGAATGTATAGAGTCTTCACAGACTTTGAGCTG  61706

Qy   1383  GATACCATGGGGTTAGGCTTCTTGACGTTCTTAGTGGATGAACAGCGGTTACCAGATTTAG  1442
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db  61707  GATACCATGGGGATTAGGCTTCTTGACGTTCTTCAGTGGATGAACAACGGTTATCAGCTGTCG  61766

Qy   1443  ATCAAGAAATTGGAAGAACTGTTTGGTTTCCCTAGTGGAAAAGGGAACCAACCCCAGGTTT  1502
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61767  ATCAAGAAGTTGGAAGAACTGTTTGGTTTCCCAAGTGGAAAATGGAACCAACCCAAGTTT  61826

Qy   1503  GACAGGGAAGAGACTTAAGGATTCGTGGGCTTCGTGGGGCTACTATTGGGACAACAATCTCACCCGCTAAACTG  1562
           |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db  61827  GAAAGGAAGAGAGCTTAAGGATTCGTGGGGCTTCGTGGGGCTACTAGTGGGACAACAATGTACCCACTAAAACTCT  61886

Qy   1563  ACGCGGTCCAAGAGCAACCAACCAAATCGGTCTACTTTCAGCGCTCGGTT  1622
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db  61887  GCGCGGTCCAAGAGCAACCAACCAAATCGGTCTACTTTCAGCGTTCAATT  61946

Qy   1623  GCCAATGTTTTTACTCCAGGAGTTCTACAGGCACCGGTCTCTAACACAGACAAATGGCAGCCATT  1682
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61947  GCCAATGTTTTTGTACTCCGAGAGTCTACAGGCACCGGTCTCTAACACAGGCAGCCATT  62006

Qy   1683  ATAGATTCAGCGCCGCTATAGGGATTCTCCGCCTTACAAAGGAAAGAATGTCCTGAGAGGA  1742
           ||||||||||||| ||||||||||| |||||||  || |||||||||  |||||||||||
Db  62007  ATAGATTCAGCTCCGCACTTCAGGGAATTCTCCGCGTACAAAGGGAAAGAATTCCTGAGAGGC  62066

Qy   1743  GATCTTAACGACTCACCACCAGTAATGCCTCGTTTGATCCATCTGTGTGGGTACGAGAAG  1802
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  62067  AATCTCAATGACGCGCCACCAGTAATGCCTCGTGTGATCCACCTGTGTGGTGATCACCTCTGTGTGTGATCACCTCTGTGATCACCAGGAAG  62126

Qy   1862  TGGGCGCTGACAAACGGCAAGAGAGGTTAAGACGCACTATGCTGGCTGGGTGGCCGTTGTG  1862
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  62127  TGGGCGCTGACAAACGGCAAGAGAGGTGAGAGACGCACTATGCTGGCTGGGTGGCCGTTGTG  62186

Qy   1863  ACGCCAATTCTGAAAGTTGTGGAGTTCCGCTCAAGGAGAGTAGGTTAGCACCGAGAATG  1922
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  62187  ACGCCAATTCTGAAGCTTCTGAGTTCCGCTCAAGGAGAGTAGCCAGGGTTAGCACCGAGAATG  62246

Qy   1982  ATGAGACTTGGATCACTTGCGCCGATGTGAGTTCTCGAGTTTGACATGGTTGGCCGACTTT  1982
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  62247  ATGGACTTTGATCATTTCGATTTCTGGGACTGTGGTTCTCGGACTATGGTTGGCCGACTTC  62306

Qy   2042  CACCCGCTACAGGTTCGAGCATTCATCGATTAGGAATCGGCCAACATTCTTTCCCTGCATT  2042
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  62307  CACCCGCTACAGGTTCGAGGATTCATCGATTAGGAATCGGCCAATATTCTTCTCCCTGCATT  62366

Qy   2102  TACGCTACTAGGATCTCGAGGGCAGGAACATTGACTTCAAGCCTCGCGTTGAAGATCTT  2102
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  62367  GACGCTAATAGGATCTCGAGGGCAGGAACATTGACTTCAAGCCTCGCGTTGAAGATCTT  62426

Qy   2162  TATTTCGAGGGCAGTCCGCCAACTGAGGAGATTAGTCACACCGAAGGAGCTACAATAGAA  2162
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  62427  TATTTTGAGGGCAGTCCGCCAACCGAGGAGATCAGTTGAGCGGAGTTTGACACGATTATCATTTC  62486

Qy   2222  GATGTTGATGAGACATATCAGATTCGACGAGACGAAGGTTGAGCGAGTTGAGCAGGAACAC  2222
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  62487  GATGTTGATGTGACAGTTGAAATAGATGAGGTGGAGTTTGACACGAGTGAATCATTATCATTTC  62546

Qy   2282  AGTGAGCATATACCTCCAGTGCAGAAGTTGGGAGAAACAGAAGTTGAGCGAGTTGAGCGAGTTGAGCAGGAACAC  2282
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  62547  AGTGAGCATATACCTCCAGTGCAGAAGTTGGGAGAAACAGAAGTTGAGCGAGTTGAGCGAGTTGAGCAGGAACAC  62606

Qy   2342  AGCAAGCTGCAGAAGTCGTGCAAGAACAACAGGATAAGTTACTCGCCAAGTGCCTCAGGGCT  2342
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  62607  AGTAAGCTGCAGAAGTGGTGCAAGAACAACTCGGATAAGTTACTCGCCAAGTGCCTCAGGGCT  62666

Qy   2402  ATCAAGTTTCTGAAGGACACAAGATCAGCTGCGCTCCTCTTCCCACTACAACTATTCCGCAATGA  2402
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  62667  ATCAAGTTTCTGAAGGACACAAGATCTTAGCTGCTCCTCTTCCTCCTCCCACTACAGCTATCCGCAGGGA  62726

Qy   2444  CAGCTCCCTCAGGACATGCCTTGACCTCCCGAGAGATATGACGGCGCCC---------  2444
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db  62727  CAGCTCCCTCAGGACATGCCTTCCCTCCCGAGAGATAATGACGGCCTAAGCCAAGTCGGCTCGGCTGTAGG  62786

Qy   2501  ---GAGCCTAGAGAGCAGAAGAGTTCTGCGAGCCATTCGTTCATCATTCGAGCCT  2501
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  62787  CCTGAGCCTAGTGAGCAGGAGAGTTCCGCATGTCCCTGCGAGGCATTCATCATTCGAGCGCT  62846

Qy   2561  CGTGAATCTAGGAGAATTGGGAGAAGAATAGGAGAACCAACCCCCAGGTTT  2561
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  62847  CGTGAATCTCGGGAGAATTGGGAGAAGAATGGGAGGAGGACCAACAAACGA  62906

Qy   2621  CTTCTCGCAGTCTCGTAGTTTACGCGCCGGTGTGGCCCGTGCCCTAAACTG  2621
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  62907  CTTCTCCAGTCTCGTAGTTTACGCGTGGCGTGCGTGTCCCAGCAGCAAACTCT  62966

Qy   2681  GAGTATCCTCAGAGCGGTGCTGGCGCGCCACAGAGCTGAGAGATCGAGTCGATGCT  2681
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  62967  GAGTATCCTATAGGCCGGTGCTGGCGCGCCACCTGCTGATGAGATCGAGTCACCACCTGCT  63026

Qy   2741  GGAGCTGATACGAACGGAACATGGCAGCAATCACACAGGCAGCCATT  2741
           |||||||||||||||||||||||||||||||||||||||||||||||
Db  63027  GGAGCTGATACAGAACAGGAACATGGCGGTTCGTGAGCAATGCAGGCAGCCATT  63086

Qy   2801  GACTACCAACTTCGTTCATTATTCGAGGTAAGCGCCTCCACTATTATTAT  2801
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  63087  GACGACCACCTCCGTTCATTCTTCGACTGAGGTAAGCGCCTCACTTCCACCATTGTTATTAT  63146

Qy   2861  ATCATTCTTGTGATTGTTCTTTATTTGTTCTTTATTTGTTTGTGATTGGATTTGTTCCTGAGTACTC  2861
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  63147  ATCATCTCTTGCGATTTGTTCTTCATTCTTCGAGTGAGGTAAGCGCCTCACTTCCACCATTGTATTAT  63206
```

```
Qy      2862 TCTTCCAAGTTTATTCACACAGTGGACTGTGTGATTAAGTTTGGGGAGGGCTCAGGAA 2921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     63207 TTTTCCAAGTTTATTCACACAGTGGACTATGTGATTTAAGTTTGGGGAGGGCTCAGGAA 63266

Qy      2922 --GTATGTTGCATTGTATATATTTTTAAGTCTGCATTCATCTAAGGCATAGAAAAACCAA 2979
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     63267 GTGTGTGTTGCATTGTTTATATTCTTGAGTCTGCATTCATCTAAGGCATAGAAAAACAAA 63326

Qy      2980 AAAAAAATTAAAAATTTCAGAAAATGATTTCACAAAAAAAGAGTGTTCAGTAGTTGCAT 3039
             |||||||| |||||||||||||||||||||||||  |||||||||||||||||||||||
Db     63327 AAAAAATTTGAAAAATTTCAGAAAAATGATTTCAC--AAAAATAGTTGTTCATGTAGTTGCAT 63384

Qy      3040 TACATTTAGGATCAAGTCTAGAGTGTTTCATTTAGGATTGTTGCATATGCATAGGGGATA 3099
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     63385 TACATTTAGGATCGAGTCTAGAGTGTTTCATTTAGGATTGTTGCATATGCATAGGGGATA 63444

Qy      3100 ATGATGAGATAGCCTTGTAAGCA 3122
             ||||||||||||||||||||||
Db     63445 ATTATGATATAGCTTTGTAAGCA 63467
```

```
RESULT 5
ATCHRIV10
LOCUS       ATCHRIV10     196296 bp    DNA     linear    PLN 16-MAR-2000
DEFINITION  Arabidopsis thaliana DNA chromosome 4, contig fragment No. 10.
ACCESSION   AL161498
VERSION     AL161498.2  GI:7267134
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 52654 to 127128; 151832 to 196296)
AUTHORS     Lamar,B., Stoneking,T., Stumpf,J., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 196296)
AUTHORS     EU Arabidopsis sequencing project.
TITLE       Direct Submission
JOURNAL     Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, John Innes Centre, Norwich, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATCHRIV9 at the 5' end and an
            overlap with ATCHRIV11 at the 3' end.
FEATURES             Location/Qualifiers
     source          1..196296
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     repeat_unit     5001..48473
                     /note="function=unclassified 22 tandemly arranged copies of
                     a 1950-bp element each element is composed of a 1890-bp
                     core flanked by 2 31-bp direct repeats element 1 resides
                     from 21901 to 23849 element 2 resides from 23850 to 25798
                     element 3 resides from 25799 to 27743 element 4 resides
                     from 27744 to 29665 element 5 resides from 29666 to 31621
                     element 6 resides from 31622 to 33582 element 7 resides
                     from 33583 to 35520 element 8 resides from 35521 to 37468
                     element 9 resides from 37469 to 39416 element 10 resides
                     from 39417 to 41364 element 11 resides from 41365 to 43313
                     element 12 resides from 43314 to 45261 element 13 resides
                     from 45262 to 47209 element 14 resides from 47210 to 49155
                     element 15 resides from 49156 to 51104 element 16 resides
```

```
                     from 51105 to 53053 element 17 resides from 53054 to 54997
                     element 18 resides from 54998 to 56946 element 19 resides
                     from 56947 to 58860 element 20 resides from 58861 to 60772
                     element 21 resides from 60773 to 62663 element 22 resides
                     from 62664 to 64578 partial element from 64579 to 65373"
     repeat_region   48761..48946
                     /note="function=centromeric repeat similar to fragment of
                     mi167, GenBank accession number X94367"
     gene            49937..52753
                     /gene="AT4g03800"
     gene            complement(join(49937..50481,50527..50943,51178..51438,
                     51481..51941,52524..52753))
                     /gene="AT4g03800"
     CDS             complement(join(49937..50481,50527..50943,51178..51438,
                     51481..51941,52524..52753))
                     /gene="AT4g03800"
                     /note="similarity to"
                     /codon_start=1
                     /protein_id="CAB80803.1"
                     /db_xref="GI:7267135"
                     /db_xref="GOA:O82612"
                     /db_xref="SPTREMBL:O82612"
                     /translation="MRSKPHQATSSEPDINRVIEEARRTLFIPQIASLRIRNSRKLNL
                     EPYNGLEDPKGYLAAFLIAAGWVDLDEAEEDARVTEHLTKDCIVLKRYLAELWASGDL
                     SKFNIDDFVKQYHEARDNSETQNSKRPRLTNEEPRSSKGKINVILGGSKLCHNSVSA
                     IKKHRPNVLLKSLSEEVNYQCSSVSFDEEETRHIERPHDDALIITLDVANFKISRIL
                     VDTGSSVDLIFLGPPSPLVAFTSESAMSLGTIKLPVLAGKMSKIVDFVVFDKPATYNI
                     ILGTPWIFQMKAVPSTYHQWLKFPTSNGVETIWGDLEGSRTSAYMPDIDPSIICHELN
                     VDPRFKPLKQKRRKLGVERAKAVNGKIDKLLKIGSIREYOYPDWVAITVVKKKNGKD
                     RVCIDFTDLNKACPKDSFPLPHIDRLVESTAGNELLTFMDAFLGYNQIMNPEDQEKT
                     SFITDRGATYQWLVNKMFNEHLRKTMEVSIDDTIVKSLKKEDHVKHLGECFEILNQYQ
                     MKLNLAKCTFGVPSGEFLGYIVTKRGIEANPNQINAFLKTPSLRNFKEVQRLTGRIAS
                     RSTDKSLPFYQILKGNNGFLWDEKCEEAFRQLKAYLTTPPVLSKPEADEKLYLYVFVS
                     NHAVSGVLVREDR"
     exon            complement(49937..50481
                     /gene="AT4g03800"
                     /number=1
     intron          complement(50482..50526)
                     /gene="AT4g03800"
                     /number=1
     exon            complement(50527..50943)
                     /gene="AT4g03800"
                     /number=2
     intron          complement(50944..51177)
                     /gene="AT4g03800"
                     /number=2
     exon            complement(51178..51438)
                     /gene="AT4g03800"
                     /number=3
     intron          complement(51439..51480)
                     /gene="AT4g03800"
                     /number=3
     exon            complement(51481..51941)
                     /gene="AT4g03800"
                     /number=4
     intron          complement(51942..52523)
                     /gene="AT4g03800"
                     /number=4
     exon            complement(52524..52753)
                     /gene="AT4g03800"
                     /number=5
     misc_feature    52654..52853
                     /note="overlap with TSH22, GenBank accession number
                     AF096372, from position 69553 to 69752"
     repeat_region   53634..54512
                     /note="function=putative transposon similar to TNP2 of A.
                     majus similar to F1K3, GenBank accession number AC006266
                     similar to T4E14, GenBank accession number AC005171"
     gene            57589..60483
                     /gene="AT4g03810"
     CDS             57589..60483
                     /gene="AT4g03810"
                     /note="strong similarity to"
```

intron   /number=1
                complement(78218..78529)
                /gene="AT4g03820"
                /number=1
exon   complement(78530..78845)

Query Match   65.0%; Score 2029; DB 8; Length 196296;
Best Local Similarity 81.3%; Pred. No. 0;
Matches 2507; Conservative 0; Mismatches 495; Indels 81; Gaps 10;

```
Qy      102  CACTTACTGGCTTCATCGCTCCATCTCCATCCTCTCAACATACTCGACCTCGCGGATATC  161
                                                                                   
Db   113057  CAAACACTCGATCTCATCGACCACTCTTCGCCACCACCTCTCATTCGTCTCCAGTTTC  113116

Qy      162  ACTCGAGCTCGCCGCTTCTCACCGCCTCTCACCGCTCACCGCCTG CTCCTCTCTCCAAG  221
                                                                                   
Db   113117  ACTCGACCACCAC---TCTCTTCGCCTCCATCATCACGGCCTGCTCCCTCTCTCCCAAG  113173

Qy      222  GAAACAACTCGAGCTCCATTCACTCACCACCTTCACCACCACCAAGCCGGCTTCACCA  281
                                                                                   
Db   113174  GAAACAACTCGACCACCTCCACCATTCATTCACTCGACCTCTCATCCACCAAGTCGCCTTCACCA  113233

Qy      282  CTTCTAGCTCTTAACCACTCGACCACCTTCACCATCAACCAATCAAATCGTTTTCTCCTC  341
                                                                                   
Db   113234  CTTCTCGCTCTTAACCACTCGACCACTCTTCACCATCAACCAATCAAATCGTTTTCTCCTC  113293

Qy      342  CATT-------AAAGCTTGACATACTCGACCGTGAACACTTATCACCTTCAAGCTCCT  393
                                                                                   
Db   113294  CATTCAAGTTCAAAACTGGACCGCGTTCACTCGACCACCAGGAAACCAGCTTCA  113353

Qy      394  CATCTCTTCATCGTTCAAACACGCGTCTCTCATCCCCACCAGAAAGCTTGTCATCACCT  453
                                                                                   
Db   113354  CTCGACTTCACAAGTTCACTCGACCGCGTCCCTTCTTCACCAATCACCGAACTCACCT  113413

Qy      454  CTCACTCATCACCAGTTCACTCGATTCAGCAACCAAACTCGACCTCCGTCTCTTCGCCAC  513
                                                                                   
Db   113414  CTCACTCTCTCATCTTACTCGACCGCGTGTAAGAAGGCTAACCAAACAATCCGATGAA  113473

Qy      514  TCATAGTCACTCGATCTCTCCTCACCATCTTCATCATCTCCCTTACTCGACCACCGTGCG  573
                                                                                   
Db   113474  CCGTCTACACACTACTCGCCCTTGCCCCGGCATCAAGTACTTGAAGGTTTCAGC----  113528

Qy      574  TCTCGGCTTCCACCATTGCCATTAAAAGCTCACTCGATTGTCAAAGAGAAGAGTGAAG  633
                                                                                   
Db   113529  -CTTGTCATTTGGGTTCGTTGGAGAATCACTCGT------CACAACATATCACTCAAC  113582

Qy      634  CTCAACCACCGCCACTCGACCGCGTTTCCCTCTACACATTCAACACTCGACCACGGTGCT  693
                                                                                   
Db   113583  CAAAGAGTTCACTCGACCACCCGGAATCTCTTTCGCTATCACCGGTTCACA--GCTCA  113640

Qy      694  ACCATCTCCACCACCGCCTCTTGTTCACCATACACTCGACCAACAACTTCAAAGTAAAAA  753
                                                                                   
Db   113641  CTCGACCGCTTCGCTTCAAAGTGTACTCGACCTCTTCCACCATCAAGCAACAACAAAACTCGA  113700

Qy      754  AAAAAAGAAAAAAAAAGTCAAAACCAAAAATCAAACCAGTTCACTCAACCGGTTACTCGACCGGTAC  813
                                                                                   
Db   113701  TTTTCAAAACCAACAAAAATCACTCGGCCGCCGGTT-ACTCAATAACATTACTTGACCAATTT  113759

Qy      814  GCTGGTTTTAGATTGTTTGTTTTTGGTTTTGCTATTACTAACATAATTTAACGTTTATCTTTGAG  873
                                                                                   
Db   113760  ACTTCTGTTAGTTTATGTTTTGTTTGGGACTAACGTTATTAACGTTTTTATCTTTGAG  113819

Qy      874  TTTCGTCTCGTTTTTTAGGTTTAGGTTTCATCATGAGAAATCCTCTATGGACCCT  933
                                                                                   
Db   113820  TTTCATCTCATTTTTCAGTTCATCATGAGAATGAGTAACTACAATGGAGAATCCCTATGGAGATCCT  113879

Qy      934  GATTATAATGTGGATGAAGCTAAGTCCTGGTCCACTAGACCGGAGTGGAGCAACATGTT  993
                                                                                   
Db   113880  GATTATAATGTAGATGAAGCTGAGTCTTGGTCCACTAGACCACTCAAAGAGAAGAGAACACGTT  113939

Qy      994  TACGAGAGCTATAGGGATGAATTTGAACGCTCGCAGCTCGAGCTAATCAAAGAGAGAGCT  1053
                                                                                   
Db   113940  TACGAAAGCTTCAAAGGAGAATTGGAACGCTCGCAGCTCGAGCTAACCAGAGAAGAAGAGCT  113999
```

```
Qy    2103 TATTTCGAGGGCAGTCCGCCAACTGAGGAGATTAGTCACCGAAGGAGAGAGCTACAATAGAA 2162
           ||||| ||||||||||||||||||  |  ||||||| | ||||||||||||  |||  || |||
Db  115080 TATTTTGAGGGCAGTCCGCCAACCAACCGAAGAGATCAGTCACAGAAGAGAGCTATAAAGGAA 115139

Qy    2163 GATGTTGATGAGACATATGATATATAGA||||| GAGCTGAGTTGACACGAGTGATCATTTC 2222
           ||||||||||| |||| || | |  |||       ||||| |||||||||||||| ||| | |
Db  115140 GATGTTGATGTGACAGTTGACAGTTGACAGTTGACAGTTGAGCTGAGTTGACACGAGTGATCATTTC 115199

Qy    2223 AGTGAGCATATACCTCCAGCGAGCGCATTGGAGACGCAAGAGTTGAGCGAGAACAAC 2282
           |||||||||||||||||||| ||||  |||||| || |||||||||||||| ||||||
Db  115200 AGTGAGCATATACCTCCAGCCAGCGAAATCAAGAGTTTGAGCGAGCTCAAAAGAACAAC 115259

Qy    2283 AGCAAGCTGCAGAGTGTGCAAGAACAGGATAAGTCACTCGCCAAGTCGCCTCAGGGCT 2342
           |||||||||||||||  | ||||||| |||||| | | ||||||| |||| |||||||
Db  115260 AGTAAGCTGCAGAGTGGTCGCAGAAAACTGATAAGTTACTCGCCAAGTCGCCTCAGGGCT 115319

Qy    2343 ATCAAGTTTCTGAAGGACAAGATCAGCTGCTGCTCCTTCCACTACAACTATTCCGCAATGA 2402
           ||||||||||||||| ||||  |||||| | |||| |||||||||||||| |||||||
Db  115320 ATCAAGTTTCTGAAGGACAAGCTTAGCTGCTGCTCCTTCCACTACAACTATCCCGCCAGGGA 115379

Qy    2403 CAGCTCCCTCAGGACATGCCTTGAGGAGGATATGACGCGCCC------- 2444
           |||||||||||||||||| ||||||||||||| ||| |||||
Db  115380 CAGCTCCCTCAGGACATGCCTTGAGGAGGATATGACGGCGGCCTAAGCCAAGTCGGCCTAGG 115439

Qy    2445 ---GAGCTAGAGACGCAGAGATTCGCATGTCCCTGCGAGCATTCATCATTCGAGCCT 2501
              ||||||| |  |||||| |||||||| ||| |||||| ||||||||||| |||||
Db  115440 CCTGAGCTAGTGACGCAGAGATTCGCATGTCCCTGCGAGGCATTCATCATTCGAGCCT 115499

Qy    2502 CGTGAATCTAGGAAGAAATAGGAGAGAGACGAAGTCGATCTGACACTCAGCAGCAGACGA 2561
           ||||||||| ||||| |||||||||||| ||||||||||| ||||||||||||||| ||||
Db  115500 CGTGAATCGGGGAGGAAGAGGAGAGAGACGAAGTCAATCTCAGCACTCAGCAGCAAAACGA 115559

Qy    2562 CTTCTGCAGTCTCGTAGTTTACGCGACCGCGGTGCGACGCCGCCACAGAGCTGCTG 2621
           |||||||||||||||||||||||||||||||||||||| || ||||||||||||||
Db  115560 CTTCTCCAGTCTCGTAGTTTACGCGACCGCGGTGTCCGAGCAGACCACCACCATGCT 115619

Qy    2622 GAGTATCCTCAGAGCCGTGCTGGCCGCCACAGAGCTGATGAGATGACTACCCACATGCT 2681
           |||| ||||| |||| ||||||||||||||||| |||| || |||| |||||| | |||
Db  115620 GAGTATTCCTCATAGCCGGTGCTGGCCGCCACAGAGCTGATGAGATCGACTACCCACCTGCT 115679

Qy    2682 GGAGCTGATACGGAACATGGGCCGGTTCGTCTATGGCCTGAGCTAACACAGGCAGCCATT 2741
           |||||||||| ||||||| ||||||||||||||| |||||||| |||||| ||||||||
Db  115680 GGAGCTGATACAGAACATGGGCCGGTTCGTCTATGGCCTGAGCTGTGACACAGGCAGCCATT 115739

Qy    2742 GACTACCAACTTCGTTCATTATTCGACTGAGGTAAGCGCCTCACTTCACCATTATATTAT 2801
           |||||||||||||||||||||| |||||||| |||||||||||||||||||||| ||||
Db  115740 GACGACCAACTCCGTTCATTATTCGACTGAGGTAAGCGCCTCACTTCACCATTGTTATTAT 115799

Qy    2802 ATCATCTCTTGTATTTGTTTCTTTATTTTGTTTCAGTGGATTGGATTGTTCCTGAGTACTC 2861
           ||||||||||| |||||||||||||| | ||| ||||| |||||||||||||||  ||||
Db  115800 ATCATCTCTTGCGATTTGTTTCTTTATTTTGTTTGGATTGGATTGTTCCTGAATACTC 115859

Qy    2862 TCTTCCAAGTTATTCACACAGTGGTGTTAAGTCTGCATTCACAAAAAAAGAGTGTAGAAAACCAA 2921
           |||||||| ||||||| |||| || || ||| |||||| |||||| |  |||  ||||||| |
Db  115860 TTTTCCAAGTTATTCACACAGTGGTGTTAAGTCTGCATTCTCGACTCTCAAGAGTAGTAGAAAAACAA 115919

Qy    2922 --GTATGTTGCATTGTATATATTTCAGAAAATGATTTTCATGTGTTCATTTGCAT 2979
             ||| |||| |||||||| | ||| || | ||| | |||||||| |||||||||
Db  115980 GTGTGTTGCATTGTATATATTTCAGAAAAAATGATCTCAAAAATAGTTGTTGCAT 116037

Qy    3040 TACATTTAGGATCAAGTCTAGAGTTGTTCATTTAGGATGTTCATATGCATAGGGGATA 3099
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  116038 TACATTTAGGATGATCGAGTCTAGAGTTGTTCATTTAGGATGTTCATATGCATAGGGGATA 116097

Qy    3100 ATGATGAGATAGCCTTGTAAGCA 3122
           ||||| | |||||  |||||||||
Db  116098 ATTATGATATAGCTTTGTAAGCA 116120
```

```
Qy    1054 GAAATCGCTAGAGAGAGAAAGAGGGCGATGTCGAGTAGATATGAGCTGATTGATGAGGATATC 1113
           |||||||| ||||||||||||||| |||||||| |||| |||||||||||||||||||| |||
Db  114000 AAAATCGCCAGAGAGAGAAAGAGAGAGCGATGTCAAGTCAAGTAGATAATGAGCTGATTGATGAGGACATC 114059

Qy    1114 AAAACTGAGTATGAGCCAGAGCGCGATGTCAAGTCGAAGCGAAGCGCTGACAACAAATCCGAC 1173
           ||||||||||||||||||||| ||||||||||| ||||||||||| |||||||||| ||||||
Db  114060 GAAACTGAGTATGAGCCAGAGCCATGGCGCCATGGCGCGAAGCGAAGCGCTGACAACAAACCCGAC 114119

Qy    1174 GAGGTTACGGTGGAGGAGTATATCAGATTCTTTGAGATGAATGACTTCTGGGGAACGAGG 1233
           ||||||||||||||||||||||||||||||||||| |||||||||||||||||||| ||||
Db  114120 GAGGTTACGGTGGAGGAGTATATCAGATTCTTTGAGATGAATGACTTCTGGGGGACGAGA 114179

Qy    1234 TATCCCTGATATGAGACTTAGCCCAGTTGGGTTACTGGAGGACGTGCAGCATCTGTTC 1293
           ||||| ||||||||||| |||||||||||||||||||||||| |||| | ||||||||
Db  114180 TATCCCTGCTATGAGACTTAGCCCAGTTGGGTTACTGGAGGACGTGCAGCATCAGTTC 114239

Qy    1294 GAGAAGTGTCATCTG--------||||| ATAAGGAGGAGACA 1322
           ||||||||||||||         ||||| ||||||||||||||
Db  114240 GAGAAGTGTCATCTGGAGACACTCGGAAATGTTACGGCTACCTTACGCTACCTATAAAGAGGAGGAGG 114299

Qy    1323 ATCGAGTTTCTTTCCACACTGCAAGTGAAATGTATGAGGGACTTCACAGACTTTGAGCTG 1382
           ||||||| ||| ||||||||||||||||| ||||||| |||||||||||||||||||||||
Db  114300 ATCGAGTTTCTTTCCACACTGCAAGTGAAATGTATGAGGGACTTCACAGACTTTGAGCTG 114359

Qy    1383 GATACCATGGGGGTTAGGCTTCTTGACGTTCTTCAGTGGATGAACAGCGGTACCGATTTAG 1442
           ||||||||||| ||||||||||||||||||||||||| ||||||| |||||||||||||| 
Db  114360 GATACCATGGGGGTTAGGGTTCTTGACGTTCTTCAGTGGATGAACAGCGTTATCAGCTGTCG 114419

Qy    1443 ATCAAGAAATTGGAAGAACTGTTTTGGTTTCCCTAGTGGAAAGGGAACCAACCCCAGGTTT 1502
           |||||||||||||||||||||||||||||| |||||||||||| |||||||||||||||||
Db  114420 ATCAAGAAGTTGGAAGAACTGTTTGGTTTCCCCAGTGGAAATGGAACCAAACCCAAGTTT 114479

Qy    1503 GACAGGGAAGAGAGCTTAAGGATTTGTGGGCTACTATTGGGAACAATGTACCGCTAAACTCG 1562
           |||||||||||||||||||| |||||||||||||||||||||||||| | || |||| |||
Db  114480 GAAAGGGAAGAGAGCTTAAGGAATTTGTGGGCTACTAGTGGGAACAATGTACCACTAAACTCT 114539

Qy    1563 ACGCGGTCAAGAGACTCACCACCACCACAAATCCGGAGTCCTGTACTTTCAGCGTCGGTT 1622
           ||||||| |||| ||||| ||||||||||||| |||||| ||||| |||||||||||
Db  114540 GCGCGGTCCAAGAGACCACCACCACCACAAATCAGGAGTCCTGTGTCTTTCAGCGTTCAATT 114599

Qy    1623 GCCAATGTTTTTACTGGGGTTAGGCTTCTACAGGCAGCCACCGTGTCTAACACAGAAGATG 1682
           |||||||||| |||| || |||| |||||||||||||| |||||| |||||||||||||||
Db  114600 GCCAATGTTTTGTACTCGTGGTTTAGGCTTCTACAGGCAGCCACCGTGTTAACGGACATGGAGATG 114659

Qy    1683 ATAGATTCAGCGCTTATAGGAGGATTCTTCCGGCCTTACAAGAAAGGAAAGATGTCCTGAGAGGA 1742
           |||||||||||| ||| ||||| ||| |||||| ||||||||||||| ||||||||||||| ||
Db  114660 ATAGATTCCGCCACTTACAGGAGGATTCTTCCGGCCTACAAAGGGAAAGATGTCCTGAGAGGC 114719

Qy    1743 GATCTTAACGACTCACCACCACCACCAGTAATGCCTCTGTTGATCCATCCATCTGTGTGGGTACATGAAG 1802
           ||| |||||| ||||||||||||||||| || |||||||||||| | | || |||||||| ||||| |||
Db  114720 AATCTCAATGACGCGGCCACCACCACCAGTAATGCCTCTGTTGATCATCCATCTGTGTGGGTGATACAGGAAG 114779

Qy    1803 ACGCCAATTCTGAAAGTTTGTGTCGCCGATGTGAGTTGTTTGCGCCCATGTCAAGTGTTGGGGGTGTTGTG 1862
           |||||||| ||||||||||| || |||||||||||||||||| ||||| |||||||||||| ||||||||
Db  114780 ACGCCAATTCTGAAAGTTGTGTCGCCGATGTGAGTTGTTGCGCCCATGTCAAGTGTTGGGGGTGGTGTG 114839

Qy    1863 ACGCCTACTACGGTTGCAGCATTCATCATCGATTCAACTGCTACTAGAATCGCCCAACATTCTTTCCCCTGCATT 1922
           |||| |||| ||| || |||||| |||||| || ||| |||||||||||| ||||| ||||| ||||| |||||
Db  114840 ACGCCTACTACGGCTGCAGCATTCATCATCGATTCATTGCTACTAGAATCGCCAACATTCTTTCTCTCTGCATT 114899

Qy    1923 ATGATCCTGGATCACTTGCGCCGCCGATGTGAGTGATCGAGTATGAGTATCGCCGCACTTC 1982
           ||||| || ||||||||||| |||| ||||||||||||||| ||| |||||| ||||||||
Db  114900 ATGGTCCTGGATCTCTTGCGCCGCCATGTGAGTGATCTCGAGTATCGAGTATCGACTAGTTGCGGCCACTTC 114959

Qy    1983 CACCGGTACAGGTTCGAGCATTCAGCATTCATCATCTTTCCCCTGCATT 2042
           |||||||| ||||||||||||||||||||||||| ||||||||||||||
Db  114960 CACCGCTACAGGTTCGAGGATTCTCGAGGCGGAGGATTCTTCCTCCCCTGCATT 115019

Qy    2043 TACGCTACTAGGATTCTCGAGTGGTTCTTGTAAGCTT 2102
           ||||||||| |||||||||||| ||| ||| ||||||
Db  115020 GACGCTACTAATAGGATTCTCGAGGAGGAAGGAACATTGACTTCAAGCCTCGCGCTTGAAGATCTT 115079
```

```
RESULT 6
T15F17/c
LOCUS       T15F17         68352 bp    DNA     linear   PLN 26-JUL-2000
DEFINITION  Arabidopsis thaliana BAC T15F17.
ACCESSION   AF262042
VERSION     AF262042.1  GI:7682788
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 68352)
  AUTHORS   Washington University Genome Sequencing Center.
  TITLE     The A. thaliana Genome Sequencing Project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 68352)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     Submitted by:
            Genome Sequencing Center
            Department of Genetics, Washington University,
            St. Louis, MO 63108, USA
            e-mail: rwilson@watson.wustl.edu

            MAPPING: Clones were assigned to the YAC map by hybridization by
            M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted
            by M. Marra, WashU, to pick the best candidates for sequencing.

            NOTICE: This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality
            >= 30); an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one m13 subclone.

NOTES:

            Coding sequences below are predicted from computer analysis, using
            the program Genefinder (P. Green and L. Hillier, ms in preparation).
FEATURES             Location/Qualifiers
     source          1..68352
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="V"
                     /map="unknown"
                     /clone="T15F17"
     gene            13450..13599
                     /gene="T15F17.j"
     CDS             13450..13599
                     /gene="T15F17.j"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="Hypothetical protein T15F17.j"
                     /protein_id="AAF67378.1"
                     /db_xref="GI:7682798"
                     /translation="MTALSYYEEKRRBRERANQEKAMIMSDKVVMSGTLEDRKLRAVE
                     KAKQP"
     gene            complement(16795..23597)
                     /gene="T15F17.i"
     CDS             complement(join(16795..18896,20555..20940,20995..22337,
```

```
                     23343..23597))
                     /gene="T15F17.i"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="Hypothetical protein T15F17.i"
                     /protein_id="AAF67377.1"
                     /db_xref="GI:7682797"
                     /translation="MAAILPVKGESTKEWVTNVIHHIESTSRPPSCIYLLSHGETDQR
                     SVRPKGDSTDNQLESPRWKPMDKTDVGPASKPKADGRKISWNKDVEEGWELVENLAQS
                     DGNYNEDCDRTVRGTADSDDKHRKDIKALNDKLDMILLSQEKHVHFLVDDEQYQVQDG
                     EGNQLEEVSYINNQSGYKGYNNFKTNNPNLSYRSTNVANPQDQVYPTQQQGQNKPF
                     VLYNQDFVPKQQFQGNYHQPPPGFAPQQNQGPAALDAEMKEMVQQLLQGQASSMEI
                     AKKLSELHHKLDCSYNNLNAKVEALNTKVRYLEGQSASTSTLKVTGLPGKSIQNPKEY
                     ATAHAITICHDRELPTQPVPDFITGDNDVQEGEASTQVEVSIVEFNHSAGSRHLTQST
                     SEEKAAILERMVKRFKPTLPSRALPWTFSKAWMERYKKSVRAKQLNEIEAVMPLMEVL
                     NLIPDPHKDVRNLILERIKLYHDSDDSDATLSRAAGKRIVQEKLEDPGSFTLPCSIG
                     EFAFSDCLCDLGASVSLMPLSVARRSRKPFGMLKDLPVMINRVEVPTDFVVLDMEVE
                     HKDPLLIGRPFLASVGAVIDVREGNIGLNLGKHIKLHFNINRTPQGSTEDGRTISRNDR
                     AISGEGETERVKELKKRSNKQDETIEKLAHTIEELRSFIMSNYSGEFSMDPDYNVDE
                     AESWSARSRECHVLQSFRDEFKRSAARRNQRRAEIARGKRAMSSRYEVINEDIETEYE
                     PESWRRETKLLNKPDBVTVEEYIRFPEMNDFWGTRYLCYETLAQLGLLEDVQHLFKKC
                     HLETLMSYPYPAYKEETIEFLSTLQVEMYEALTDFELDIMGLGFLTFLVNEQRYQLSI
                     KKLELFGFPSGKGTKPRFDREELKDLWATIENNLPLNLARSKSNQIRSHVICYFQRS
                     VANVFYSRESTGTVFNTDIEMIDYALTGILRRTKGKNVLRGDLNNAPQVMPLLIHFCG
                     YRKWALINGKKKVRGALCVGGVVTPILQGCGVFKREGLAPRMMDLDHLRRCEFLEFD
                     MAGDFHRYRFPDNSSIKIANILLPCIDATRILEGRNIDFKPALEELYFEGSPPTEEISH
                     TEGATTKDIDETDNIDEAEFDTSMYHFSEHIPPARESRSLSEAHRNNSKLQKWCMKQD
                     KLLAKCHRAIKFMKNKLJSCYSSNTAIPQGHLPQDMSRRYDAPEPSRRRHEPREQEIP
                     HVPARHSSFEPRESGRKRRTLTRSSRSGRLLHSRSLRDRGAGRSRRREVEYPQSGA
                     GRHRADEVEYPPAGADTEQGGSLITSNYVHSSTEVSASLHHCIIPSIVICFFIFFISV
                     IGFVLSTLFQVYSHSGLCDLSLGEGSGSVCCIVHIFESAFI"
     gene            complement(24606..24944)
                     /gene="T15F17.h"
     CDS             complement(24606..24944)
                     /gene="T15F17.h"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="Hypothetical protein T15F17.h"
                     /protein_id="AAF67376.1"
                     /db_xref="GI:7682796"
                     /translation="MAEEONQQNGPANIGVGDAPDHRQSKGIEPPAIQNNFEIKSG
                     LISMIQGNKFHGLPMEDPLDHLDEFDRHYNLKKINGVSEDGFKLCLFPFSLGNKAHIW
                     EKNCNARVPS"
     gene            29265..30290
                     /gene="T15F17.k"
     CDS             29265..30290
                     /gene="T15F17.k"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="Hypothetical protein T15F17.k"
                     /protein_id="AAF67379.1"
                     /db_xref="GI:7682799"
                     /translation="MSKLTNLMYAALLAFGDNYLQWALDTKIELKSKGLARCIVEGNE
                     KTERSKYQAISIIRRHIAESLKNQYLTVEDPLEWLELKNRYDHQRTIQLPKAQHDWL
                     NLRIQDYKSVEEYNSELFKIVSILRLCGEKVTENDMLEKTFSTFHANNVLLQQQYRAK
                     GFTTYTSLASCLLAEKNNELLLMNSALRPPGSTAVPEANRAEMAKAPNEPQATKESN
                     YVHRGNPHGRGHGRGRGRGRGNFYGQGNHYGGRGRGINKPRGKAKS
                     VCYRCGMDDHWAKTCRTSKHLIEAYQEMIKQKGPEANLVHLDGEGDFDHEKDDLMDYE
                     TSDILGE"
     gene            30441..35665
                     /gene="T15F17.l"
     CDS             join(30441..31166,31218..33390,33445..33789,35484..35665)
                     /gene="T15F17.l"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="Hypothetical protein T15F17.l"
                     /protein_id="AAF67380.1"
                     /db_xref="GI:7682800"
                     /translation="MAEDKDILIVDSGSSHTILRDKRYFINLTLRNANISTIAGIASL
                     IEGYGQAHVLLPNGTHLEISDALYSPSKRSLLSFKDIRLNGFHVETKGEGNREFLHI
                     TEIAQGHKRVLETIPALSTGLVHTKINMIEANLAMNKEFIEEFTLWHDRLGHGPGHNMM
                     RKLMISSKGHTLKEKRVIPKKJTCAACAQGKLIIRPSPAKVNKETINFLERIQGDICG
                     PIHPPCGTFRYFMVLIDASTRWSHIIRLRAHPDFPLKTIRLDNAGEFTSQAFNDYCM
```

```
Qy   343 ATTAAAGCTTGACATACTCGACGCTGAACACTTATCACCTTCAAGCTCCTCCATCTCTTC 402
             |||||  ||||| |||||||||||||||          |||||||||||||||||||||
Db 48693 ATTACAGCTTC---ACTTGACATACTCGACGCCGCGCTCATCATTATCATCTTTAAGCTCCTCCATCTCTCC 48638

Qy   403 ATCGTTTCCAACACCGCTGCTCTCATCCCCCAGGAAAGCTTGTCATCATCTTCCTTCACTCAT 462
             |||||||||||||||||||||||||||||||  ||||||||||||||||||||| |||||||
Db 48637 ATCGTTTCCAACGCCGCCGCGCTCTCAT---CCCACGAAAGCTCGTCATCATCTTTCACTCAA 48580

Qy   463 CACCAGTTCACTCGATTCAGCAACCAAACTCGACCTCGTCTCTTGCCACTCATAGTCA 522
             |||| |||||||||||||||||||||||  ||||||||||||||||  || ||||||
Db 48579 CAACGGTTTACTCGATTCAGCAACCATAC---ACTCGATTCCTCATCACTTCTTCGTCC 48523

Qy   523 CTCGATCTCCTCCACCATCTTCATCATCTCCCTTACTCGACCACCGTGGTCTCGCTCC 582
             |||||||||| |||||||||||||||||||||||| |   |||||||||||||||||
Db 48522 AAGCATCACTACTCGACCT-TGCCTAATCACTCGACCT---CACCATCTATCTTCGTCT 48468

Qy   583 ACCATTGCCATTTAAAAGCTCACTCGATTGTCAAGAGAAGAAGAGTGAAGCTCAACCAC 642
             |||||||  ||||||||||||||||||||||||||||  |||| |||||| |||||||
Db 48467 CTCATTTGCTCAACAACAGTCATTTCTCCTCTCCACACTCGACCTCACCTCAACACA 48408

Qy   643 CGCCACTCGACCGCGTTCCCTCTCACACATTCAACACTCGACCACGGTGCTACCACTCTCC 702
             |||||||||||||||||||||  |||||||||||||||||||||||  |||||||||||||
Db 48407 GCTTTTCTTAGCTAGTCCACTCCATAACTCCACTCAACTTCACGGCCTTCGTTCTCACC 48348

Qy   703 AC--------ACCCGTCTTGTTCACCATACACTCGACCAACAACAACTCAAAGTAAAA 752
             ||        ||||||||||||||||||||||||||||||| |||||||||  |||| |
Db 48347 ACCGATTCAACAACTCGAATCACCGTCGTCTCCACTCGGCCTTCCGAAAACACTC 48288

Qy   753 AAAAAAAGAGAAAAAAAGTCAAAACCGAATTGGTACACTCGACCGGTA 812
             ||||||||    |||||||| ||||||||||||  |||||||||||||
Db 48287 AACCCAAGTCCACAAACCGAAAAAAAAAAATTGGTACACTCGACCGGTT 48228

Qy   813 CGCTCGTTGTAGATTGTGTGTTTTTGGTTTTGCTATTAACATATTAACGTTTATCTTTGA 872
             ||||| ||||||||||  ||||||||||||||  |||  |||| ||||||||||  ||||
Db 48227 TACTCTTTTGGTTTGGTTTGTGTTTTTGGTTTTGTGGAACTAACGTTATTAACGTTTATTTTTGA 48168

Qy   873 GTTTCGTCGTCGTTTTTAGGTTTTCATCATGAGTAACTACAGTGGAAAATCCTCTATGGACCC 932
             ||||||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 48167 GTTACATTTGTTTTGTGTTTCATCATGAGTAACTACAGTGGAGAATCCTCTATGGACCCC 48108

Qy   933 TGATTATAATGTGGATGATGAAGGTAAGGTCCTGGTCCCACTAGAGCCGGAGTGGAGCAACATGT 992
             ||||   ||||||||||||||||||||  ||||||||||||||||||       |||||||||||
Db 48107 TGACTACAAATGTGGATGAAGTCGAATCCTCGGTTCCGTCTAGACC---GAGAGCGAACAACAATGT 48051

Qy   993 TTACGGAGCTATAGGGATGAATTGAACGCTCGACGCTCGCAGTAATCAAAGAAGAGC 1052
             ||  |||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db 48050 TTTCCAGAGCTTCAGGGATGAATTGAACGCTCGACGCTCGCAGTAATCAGAAGAAGAGC 47991

Qy  1053 TGAAATGCTAGGAGGAAAGAGGGCGATGTCGAGTAGTAGATAGGAGAGGATAT 1112
             ||||| ||||||||||||||||||||||||||| ||||||| ||||||||||
Db 47990 TGAAATCGCTAGGAGGAAAGAGACGCGATGTCAAGAGATAGAGCTGATTGATGAGGAAAT 47931

Qy  1113 CAAAACTGAGTATGAGCCGAAGAGCAAGGAGCAAGTGCTGAACAAATCCGA 1172
             |||||||||||||||||||||||||||||||||||||||||||   |||||
Db 47930 CGAGACTGAGTATGAGCCAGGAGCAGGGAGCAGGGAGCAAGTGCTGAACAAACCCGA 47871

Qy  1173 CGAGGTTACAGTGGAGTGGAGGTATATCAGATTCTTTGAGGTGAAGTGAATGAACCAG 1232
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 47870 CGAGGTTACAGTGGAGTGGAGGTATATCAGGTTCTTTGAGGTGAAGTGAATGAACCAG 47811

Qy  1233 GTATCCCTGATATGAGACTTTAGCCCAGTTGGGGTTACTGGACGTGCAGCATCTGTT 1292
             ||||||| || ||||||||||||||||||||||||||||||||||||||||||||||
Db 47810 GTATCCGTGCTATGAGACTTTAGCCCAGCTGGGGTTACTGGACGTACAACATCTGTT 47751

Qy  1293 CGAGAAGTGTCAT------------------------------------CTGATAAGGAGGAGAC 1321
             |||||||||| ||                                    |||||||||||||||
Db 47750 CGAGAAGGGTCATATGGAGACACTCATGTCTTATCCCTTATCCGTCTTATAAGGAGGAGAC 47691

Qy  1322 AATGGAGTTCTTTCCCACTCGAAATCGTTTTCTCCTCC 1381
             |||||||| ||||||||||||||||||||| |||||||
Db 47690 AACCGAGTTCTTTCCCACCGTCGCAAGTGGAGATGTGATGAGAGCACTTGAGCT 47631

Qy  1382 GGATACCCATGGGGTTAGGCTTCTTAGTGATGAACAGCAGCGTACCAGATTTA 1441
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 48724 CACCAATCAACCAACCAAATTGA-TTCTCCTCC 48694
```

SMGVSVEHPVAHVHTQNGLAESFIKRIQLIARPLLMRSRLPVAAWGHAVLHASELIRI
RPSSEHKYSPSQLITGHEPDISHLRTFGCAVYVPIAPPQRTKMGPQRRMGIYVGFDSP
TIIKYLEPTTGDLFKARYADCHFSESEFPALGGETNKLGKEIKEIAWNQTSLNWQDPQ
TLMCDSEVHKIIHLQKLANELPDSFADPKRVVKSYIPACNAQVRIDIQKGINVIATES
NPQKKRGRPVGSKDKIPRKTKKGAMGTEVKETIDMAAASPKEPPNEIRDAEAHGPEGI
DNDEISINYIMSGIKWNRKEVDVDEKFAYEVAYEINEDHEPTSIIECTQRSDWLKWKE
AIDVELGSLKKRDVFGPILRTPSNIKPVGHKWVFVRKRNEKNEIVRHKARLVAQGFSQ
RPGVDYEETYSPVVDATTFRYLISLAIRENLDLRLMDVVTAYLYGPLDNEIYMRLPEG
IELKGKDNGSRDQYCIRMNRSLYGLKQSGRMWYNRLSEYLVKEGYKNDPISPCIPIK
KFASKGFVIIAVYVDDLNILGTSGEIAQTVEYLKKFEMKDLGKTKFCLGLQLEYVDK
GILVHQRAYTETILKRFNMDXAHPLTSPMQVRSLGLDSDPFGPKKDDEEILGPEMPYL
SAIGAWMYLSSHTRPDICFAVNLLSRFSSCPTKRHWEGIKHLLRYLQGTIDFGYLSDP
HNGKSQTGYVFTHGGTAISWRSMKQTIAATSSNHAEILAMHEASRECWLRSMTQHIR
ADCGMVEDKEPTIIYEDNAACIAQLKEGYIKGDRTKHILPKFFTHDLQKDVCEPDAA
AVCEPDAAAFPDATAFSDPANLPTDLADLPTDAAAFNGDFPLVAEV"

gene   37905..38945
            /gene="T15F17.m"
CDS   37905..38945
            /gene="T15F17.m"
            /codon_start=1
            /evidence=not_experimental
            /product="Hypothetical protein T15F17.m"
            /protein_id="AAF67381.1"
            /db_xref="GI:7682801"
            /translation="MADVVDEQEPTNIGAGDFPHHNQNRHGIVPPVQNNNFKIKSS
LIAMVQGNKFHGLPMEDPLDHLDEFERLCDLTKINGVSEDGFKLRLFPFSLGDKAHLW
EKTLPQGSITTWDDCKKVFLEKFFSNSRTARLMNEISGFTQKQNESFCEAWERFKGYQ
TKCPHHGFKEASLLSTLYRGVLFKIRMLLDTASNGNFLNKDVEDGWEIVENLAQSDGN
YNEDYDRSIPTSTESDDKHRREIKALHEKIDKLLQVQQKHVHSASEDEVFQIQEGEND
QDABISYVQNQGGYNKGYNNYRPNPNLTYRSTNVANPQNQVYPQQHSRTKPNRLFLTT
KAKGSFLLIIGF"

gene   complement(42508..43164)
            /gene="T15F17.g"
CDS   complement(42508..43164)
            /gene="T15F17.g"
            /codon_start=1
            /evidence=not_experimental
            /product="Hypothetical protein T15F17.g"
            /protein_id="AAF67375.1"
            /db_xref="GI:7682795"
            /translation="MVDDFHKPRFEHSTDRRANILLPSPEVTRIIEGDNIDFRPEVGR
LYYENAPPIDEDEFLEEAASDGMDEDRAVEFDTSMYHFGEHVPPARQSKSLTKPHKNN
SKLQKWCKHDKLIAKCFKLQTDKLSCSSSTTAIPVQOPPLEMPSRRFNAPAHRPELS
EQRVPHVQARHSFESREHKRRRKATLTRSSRSRLIHSRRSLDCGAGRSRRREVEYP
"

gene   complement(46220..47737)
            /gene="T15F17.f"
CDS   complement(46220..47737)
            /gene="T15F17.f"
            /codon_start=1
            /evidence=not_experimental
            /product="Hypothetical protein T15F17.f"
            /protein_id="AAF67374.1"
            /db_xref="GI:7682794"
            /translation="METLMSYPYPAYKEETTEFLSTLQVEMYEALTDFELDTMGLGFL
TFSVNEQRYELSIKKLEELFGFPSGKGTKPFDREELKDLWATIGNNLPLNSARSKSN

Query Match   63.2%; Score 1973; DB 8; Length 68352;
Best Local Similarity 82.4%;
Matches 2498;   Conservative   0;   Mismatches   395;   Indels 138;   Gaps 15;

Qy   163 CTCGAGCTCGGCCGTCGCCCGTTCTCACCGGCCTCTCCATCGTCACCGCCTCCTCCTGCTCCTCGTTT 222
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 48823 CACTCGATCGCCGCCCGTTCTCACCGGCGGTTCACCAC 48764

Qy   223 AAACAACTCGAGCTCTCCATTTCACTCACTCGACCTCTC--------- 282
             ||||||||||||||||||||||||||||||||||||||
Db 48763 AAACAACTCGAGCTCTCCATTTCACTCACTCGACCTCTC---------- 48725

Qy   283 TTCTTAGCTCTTAACCACTCGACCACTTCCACCATCAACCAAATGTTTTCTCCTCC 342
             |||||||||||||||||||||||||||||||||||| || ||| ||||||||||
Db 48724 CACCAATCAACCAACCAAATTGA-TTCTCCTCC 48694
```

```
Db  46550 GCATGAGCCTTAGAGGAGCAGGAGAGATTCCGCATGTCCCTGCGGAGGCATTCATCATTGAGCC 46491

Qy   2501 TCGTGAATCTAGGAGAACATAGGAGGAACGACACTCACTCGATCTGATCTCAGCAGCAGGAGCAGACG 2560
           ||||  ||| |||||||||||||| ||||||| ||| |||||||||||||||||||||||||||||
Db  46490 TCGTAAATCAGAGGAGGACGAGGAGGACCACCACTCGCTCGATCTCAGCACTAGGAGCGGACG 46431

Qy   2561 ACTTCTGCAGTCTCGTAGTTTACGCGGACCGGTGCTGGCGCAATAGAAGAAGAGAGGAGAGT 2620
           |||||||||||||||||||||||||||  |||||||||||||||||||||||||||||||||
Db  46430 ACTTCTGCAGTCTCGTAGTTTACACGCGGACCGGTGCTGGCGCCAGCAGAGGAGGAGAGGT 46371

Qy   2621 CGAGTATCCTCAGAGCGGTGCTGCGCGCCACAGAGCTGATGAGATCGAGTACCCACATGC 2680
           |||||||||||||||||||||||||  |||||||||||||||||||||||||||||||||
Db  46370 CGAGTATCCTCAGAGCGGTGCTGCGCGCCACAGAGCTGATGAGGTCGAGTACCCACCTGC 46311

Qy   2681 TGGAGCTGATACGGAACATGGCGGTTCGTTCGTCTCATCGGCTTGGGAGCAATCACAGGCAGCCAT 2740
           |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  46310 TGGAGCTGATACAGAACAAGGTGTTCATCTCATCGGCGGAGCAATCACAGGCAGCCAT 46251

Qy   2741 TGACTACCAACTTCGTTCATTATTCGACTGAGGTAAGGCCCTCACTTCACCATTATATTA 2800
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||  |||
Db  46250 TGACGAGCAAACACGTTCATTCATTCGACTGAGGTAAGTGCCCTCACTTTACCATTGTATTA 46191

Qy   2801 TATCATCTCTTGTGATTTGTT----CTTTATTTTGTTTCAGTGATTGGATTTGTCCTGA 2855
           |||||||||||||||||||||    ||| || ||||||||||||| ||| ||||||||||
Db  46190 TACCATCTATTGTGATTTGTTTCTTCATTTTTTGTTTCTATGATTGGATTTGTCCTGA 46131

Qy   2856 GTACTCTCTTCCAAGTTTATTCACACAGTGGACTGTGTGATTTTAAGTTTGGGGAGGGCT 2915
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  46130 GTACTCTCTTCCATGTTTATTGACACGTGGACTGTGTGATTTTTAAGTTTGGGGAGGGCT 46071

Qy   2916 CAGGAAGT--ATGTTGCATTGTATATATTTTTAAGTCTGCATTCATCTAAGGCATAGAAA 2973
           ||||||||  |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  46070 CAGGAAGTGTGTGTTGCATTGTGTATATTTTTGAGTCTGCATTCATCTGAAGGCATAGAAA 46011

Qy   2974 AACCAAAAAAAAATT-AAAAATTTCAGAAAATGATTTCAC-AAAAAAAGAGTTGTTCATGT 3031
           |||||||||||| | ||||||||||||||||||||||||| ||||||||||||||||||||
Db  46010 AACCAAAAAAAATTGAAAATTTTCGGAAAATGATTTCACAAAAAAAAGAGTTGTTCATGT 45951

Qy   3032 AGTTGCATTCATTAGGATCAAGTCTAGAGTGTTTCATTGGATTGTTGCATAATGCAT 3091
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  45950 AGTTGCATTACACATTAGGATCGAGTCGTTCAGAGTGTTTCATTTAGGATTGTTGCATATGCAT 45891

Qy   3092 AGGGGATAATGATGAGATAGCCTTGTAAGCA 3122
           |||||||||||||||||||||||||||||||
Db  45890 AGGGGATGATGAGATAGCCTTGTAAGCA 45860
```

RESULT 7
```
AB046433/c
LOCUS       AB046433        116966 bp    DNA        linear   PLN 18-APR-2001
DEFINITION  Arabidopsis thaliana DNA, chromosome 5 centromere region,
            clone:F11C12.
ACCESSION   AB046433
VERSION     AB046433.1  GI:9971604
KEYWORDS    centromere.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (sites)
  AUTHORS   Kumekawa,N., Hosouchi,T., Tsuruoka,H. and Kotani,H.
  TITLE     The size and sequence organization of the centromeric region of
            arabidopsis thaliana chromosome 5
  JOURNAL   DNA Res. 7 (6), 315-321 (2000)
  MEDLINE   21082928
   PUBMED   11214966
REFERENCE   2  (bases 1 to 116966)
  AUTHORS   Kotani,H. and Kumekawa,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUL-2000) Hirokazu Kotani, Kazusa DNA Research
            Institute, Lab. Chromosome II; 1532-3 Yana, Kisarazu,
            Chiba 292-0812, Japan (E-mail:kotani@kazusa.or.jp,
```

```
Db  47630 GGATACCATGGGGGTTAGGCTTCTTCTTGACGTTGTTCTCCAGTGAATGAACAGCAGCCGTATGAGCTTTC 47571

Qy   1442 GATCAAGAAATTGGAAGAACTGTTTGGTTCCCTAGTGGAGGAACCAACCACCCAGTT 1501
           ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  47570 GATCAAGAAGTTAGAGAGAACTGTTTGGTTTCCCTAGTGGAGGAACCAAACCCAGATT 47511

Qy   1502 TGACAGGGAGAAGCTTAAGAGCTTGTGGGCTACTATTGGGACAACAATCTACCGCTAAACTC 1561
           ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db  47510 TGACAGAGAGGAGCTTAAGGATTGTGGGCCACTATAGGGACAACAATCTACCGCTAAACTC 47451

Qy   1562 GACGCGGTCCAAGAGCAACCAAATCCGAGTCCTGTGATTCGCTACTTTCAGCGCTCGGT 1621
           ||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db  47450 GGCGCCGGTCTAAGAGCAACCAAATCCGAGTCCTGTGATCCGCTACTTTCAACGATCGGT 47391

Qy   1622 TGCCAATGTTTTTTACTTCCAGGGAGTCTACAGGCACCGTGTCTAACACAGACATGAGAGAT 1681
           |||||||||| ||||||||||||||||||||||||||| |||||||||||||||||||||||
Db  47390 TGCCAACGTCTTTTACTTCCAGGGAGTCTACAGGCACCGTGTCTAACACAGACATGGAGAT 47331

Qy   1682 GATAGATTCAGCGCTTATAGGGATTCTCCGCCTTACAAAAGGAAAGAATGTCCTGAGAGG 1741
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  47330 GATAGATTATGCGCTTACAGGGATTCTCCGCCGGACGAAGGGGAAGAATGTTCTGAGAGG 47271

Qy   1742 AGATCTTAACGACTCACCACCAGTAATGCCTCTGTTGATCCATCTGTGGTACATGAA 1801
           ||||||||| ||||||||||||||||||||||||||||| |||||||||||||||||
Db  47270 AGATCTCAACACGACCACCAGTAATGCCTTCTGTTGATCCATCTGTGGGATACAGGAA 47211

Qy   1802 GTGGGGCTGACAAACGGCAAGAGGGTAAGAGAGCACTATGCGTGGCGTTGGCGTTGT 1861
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  47210 GTGGGGCCGTTGACAACGGCAAGAGGGAGAGAGAGGTGAGGAGGACTATGCGTGGCGTTGT 47151

Qy   1862 GACGGCCAATTCGAAAGTTCTGGAGTTCCGTCAAGGAAGGTAGGGTTAGCACCGAGAAT 1921
           ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db  47150 GACGGCCAATTCGAAAGTTCTGGAGTTCCGTCAAGGAAGGTAGGGTTGGCACCGAGGAT 47091

Qy   1922 GATGGACTTGGATCACTTGCGCCGATGTGAGTTCTCTGAGTTGACATGGTTGGCGACTT 1981
           ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db  47090 GATGGACTTGGATCACTTGCGCCGATGTGAGTTCTCTGGAGTTGACATGGTTGGCGACTT 47031

Qy   1982 TCACCGCTACAGGTTCGAGCATTCATCGATTAGAATCGCCAACATTCTTTTCCCCTGCAT 2041
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db  47030 TCACCACTACAGGTTCGAGCATTCATCGATTAAAATCGCCAACATTCTTCTCTCCCTGCAT 46971

Qy   2042 TTACGCTACTAGGATTCTCGAGGGCCAGGGCGAGGAACATTGACTTCAAGCCTGCGCTTCAAGAGATCT 2101
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db  46970 TGACGCTACTAGGATTCTCGAGGGCCAGGGACGAACCATTGACTTCAAGCCTGCGCTTGAAGAACT 46911

Qy   2102 TTATTTCGAGGGCAGTCCGCCAACTGAGGGAGATTAGTCACACCGAAGGAGCTACAATAGA 2161
           ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db  46910 TTATTTCGAGGGCTAGTCCGCCAACTGAGGGAGATCAGTTTTACCGAAGGAGCTACAACAA 46851

Qy   2162 AGATGTTGATGAGACATATGATATAGATGAGGCGGAGTTTGACACGAGTTGACACGAGTGATCATCTT 2221
           |||||||||||||| |||| |||||||||||||||||||||||||||||||||||||||||||
Db  46850 AGATATTGATGAGACTGATGATATAGATGAGGCTGAGTTTGACACGAGTGATCATTT 46791

Qy   2222 CAGTGAGCATATACCTCCGAGGGAAAAGCAAGAGTTTGAGCGAAGCTCACAGGAACAA 2281
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db  46790 CAGTGAGCATATACCTCCTTGCGAGGGAAAAGCAAGAGTTTGAGCGAAGCTCACATGAACAA 46731

Qy   2282 CAGCAAGTTGCAGAAGTGGTGCAAGAAACAGGATAAGTTACTCCGCCAAGTGCCTCAGGGC 2341
           ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db  46730 CAGCAAGTTGCAGAAGTGGTGCAATGGTGCAAGAAACAGGACAAGTTACTCCGCCAAGTGCCTCAGGGC 46671

Qy   2342 TATCAAGTTTCTGAAGGACAAGATCAGCTGCTCCTCCTCCACTACAACTATTCCGCAATG 2401
           ||||||||| ||||||| | |||||||||||  | |||||||||  |||||||||||||||
Db  46670 TATCAAGTTTATGAGAGAACAAGCTGCAGTCGTCGACCCTTCTTCACTATAGCTATTCATTTCCGCAGG 46611

Qy   2402 ACAGCTCCCTCAGACACATGCCTTCGAGGAGACATATGACCGGCCC---------- 2444
           ||| |||||||||||||||||||||||||||||||||| ||||||||
Db  46610 ACATCTCCCTCAGGACACATGCCTTCGAGGAGACATATGACGCGCCCGAGCCAAGTCGGCGCAG 46551

Qy   2445 ---GAGCCTTAGAGAGCAGAAGATTCTGCATGTCCCTGCGGAGGCATTCATCATTCGAGCC 2500
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
FEATURES          Tel:81-438-52-3920, Fax:81-438-52-3921)
        source    Location/Qualifiers
                  1..116966
                  /organism="Arabidopsis thaliana"
                  /mol_type="genomic DNA"
                  /cultivar="Columbia"
                  /db_xref="taxon:3702"
                  /chromosome="5"
                  /map="centromere"
                  /clone="F11C12"
        misc_feature   1..3000
                  /note="overlap to BAC F22D12, please refer for analysis
                  and annotation"
        LTR       3169..3588
        misc_feature   3589..5678
                  /note="similarity to fragments of Tat retrotransposons"
        LTR       5679..6977
        misc_feature   6978..14178
                  /note="similarity to fragments of Athila retrotransposons"
        LTR       14179..16201
        misc_feature   16202..16868
                  /note="similarity to fragments of Athila retrotransposons"
        LTR       16869..18727
        misc_feature   18728..19645
                  /note="similarity to fragments of Tmal retrotransposons"
        LTR       19646..19982
        misc_feature   19983..25739
                  /note="similarity to fragments of copia type
                  retrotransposons"
        LTR       25740..26076
        misc_feature   26077..27647
                  /note="similarity to fragments of En/Spm transposon"
        LTR       27648..29275
        misc_feature   29276..30587
                  /note="similarity to fragments of Athila retrotransposons"
        LTR       30588..32377
        misc_feature   32378..34900
                  /note="similarity to fragments of Athila retrotransposons"
        LTR       34901..36228
        misc_feature   36229..43996
                  /note="similarity to fragments of Athila retrotransposons"
        LTR       43997..45230
        LTR       45231..47023
        misc_feature   47024..54888
                  /note="similarity to fragments of Athila retrotransposons"
        LTR       54889..56774
        LTR       56775..57057
        misc_feature   57058..57674
                  /note="similarity to fragments of Athila retrotransposons"
        LTR       57675..59455
        misc_feature   59456..64584
                  /note="similarity to fragments of Athila retrotransposons"
        LTR       64585..66360
        misc_feature   66361..74716
                  /note="similarity to fragments of Athila retrotransposons"
        LTR       74717..76300
        misc_feature   76301..83804
                  /note="similarity to fragments of En/Spm transposon"
        misc_feature   83805..87751
                  /note="similarity to fragments of Tmal retrotransposons"
        LTR       87752..88909
        LTR       90385..91966
        misc_feature   91967..94823
                  /note="similarity to fragments of Athila retrotransposons"
        LTR       94824..95552
        LTR       95553..99722
        misc_feature   99723..100331
                  /note="similarity to fragments of Tat retrotransposons"
        LTR       100655..101120
        LTR       101121..102789
        misc_feature   102790..104802
                  /note="similarity to fragments of Tat retrotransposons"
        LTR       104803..105528
        misc_feature   105529..113495
                  /note="similarity to fragments of Tat retrotransposons"
        misc_feature   113496..116530
                  /note="similarity to fragments of En/Spm transposon"
        LTR       116531..116966
                  /note="similarity to fragments of Tat retrotransposons"
ORIGIN
```

```
Query Match               63.0%;  Score 1967.2;  DB 8;  Length 116966;
Best Local Similarity     81.7%;  Pred. No. 0;
Matches 2468;  Conservative  0;  Mismatches  433;  Indels  121;  Gaps  12;

Qy    163 CTCGAGCTCGCGCTTCCACCGCCTCTCATCGTCACCGCCTGCTCCCTCTCTCCAAGG 222
          ||||||||| | || ||||| |||| |||||| ||||||||||||||||||||||||
Db  40481 CACTCGATCGCCGGCTTCTCACCGCCTCTCCATCATCGTTGCCTACTCACTCTCCAAGG 40422

Qy    223 AAACAACTCGAGCTCCATTTCACTCACTCGACCTCC------------------ 282
          ||||||||||||||||||||||||||||||||||||
Db  40421 AAACAACTCGAGCTCCATTTCACTCACTCGACCTCTC----------------- 40383

Qy    283 TTCTAGCTCTTAACCACTCGACCTTCACCATCAACCAATCAAATCGTTTCCTCCTCC 342
          |||||| |||  ||||||| |||||||||||||||| |-||||||||||||||||||
Db  40382 -----------------------------CACCATCAACCAACCAAATTGA-TTCTCCTCC 40352

Qy    343 ATTAAAGCTTGACATACTCGACCGTCTGAACACTTATCACCTTCAGGCTCCTCATCTCTTC 402
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  40351 ----------ACTCGACGCCGCTGCTCCAT---CCCACGAAGCTCGTCATCTTCCTTCACTCAA 40296

Qy    403 ATCGTTTCCAACACCGCTGCTCTCCATCCCCCACGAAAGCTTGCTCATCACCTCCACTCAT 462
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  40295 ATCGTTTCCAACGCCGCTGCTCCAT---CCCACGAAAGCTCGTCATCTTCCTTCACTCAA 40238

Qy    463 CACCAGTTCACTGATTCAGCAACCAACCAAACTCGACCTCGTCTCTTCGCCCACTCATAGTCA 522
          ||||||||||||||||||||||| ||||||||||| ||    ||||| |||||||||||||
Db  40237 CAACGGTTTACTCGATTCAGCAACCATCACTCGATTCCTCATCTTCGTCGTCCAAG 40178

Qy    523 CTCGATCTCTCCTCCACCATCTTCATCATCTCCCTTACTCGACCACCGTCGCTCGCTCC 582
          |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db  40177 CATCACTACTCGACCTTGCCTAATCACTCGACCTCACCATCTATCTTCGTCTCTCATTTG 40118

Qy    583 ACCATTGCCATTTAAAAGCTCACTCGATTGTCAAAGAAGAAGAGTGAAGCTCAACCAC 642
          |||||| |||||||||||||||| ||||||||||||| |||||||||||||||||||||
Db  40117 CTCAACAAACAGTCATTTCTCCTCTCCAAGGAATACACTCGACCTCAACACAGTTTCCT 40058

Qy    643 CGCCACTCGACCGCGTTCCCTACACATTCAACACTCGACCACGTGCTACCATCTCC 702
          || || |||||||||||| ||||||||||||||| ||||||||||||||||| ||
Db  40057 AGCTCATAACTCCACTCACTTCACGCGCCAAGAGCACCTCCGTTTCTCAACCGATTCA 39998

Qy    703 AC-ACCCGCTCTTGTTCACCATCAGTTCACTCGACCAACTCTCAAAGTAAAAAAAAAGA 761
          || ||||||||| || ||| ||||||| ||| ||||||||||| |||||||||||||||
Db  39997 ACAACTCGAATCACCCGCTTCGTCTCACCGGCCTTCCGAAAACACTCACCGCAAG 39938

Qy    762 AAAAAAAGTCAAAACCGACAGTTTCACTCAACCGGTTTCACTCGACCTACGCGTGGTTT 821
          ||||||||||||||||| || ||||||||||| || ||||||||||||| |||||||||
Db  39937 TTCACAAAACCGAAAAACAAAAAAAAAATTGGTACACTCGACCTCACCGGTTTACTCTTTT 39878

Qy    822 AGATTGTGTTTTGGTTTTGCTATTACTAAACATATTAACGTTTATCTTTGAGTTTCGTCT 881
          ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db  39877 GGTTTGTGTTTTTGGTTTTGTTGGAACTAACGTATTAACGTTTATTTTTGAGTTTACATTT 39818

Qy    882 GTTTTTAGGTTTCATCATGATGGAGTAACTACAGTCGAAAATCCTCATGGACCCTGATTATAA 941
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  39817 GTTTTTCAGTTTCATCATGATGGAGTAACTACAGTACAGTGGAGAATCCTCATGGACCCTGACTACAA 39758

Qy    942 TGTGGATGAAGCTAAGTCCGTCGGTCCACTAGACCGGAGTGAGAGCAACATGTTTACGAGAG 1001
          |||||||||||||||||||||||||||       |||||||||| || ||||||||||||
Db  39757 TGTGGATGAAGCTGAATCCGGTCCGCTAGACC---GAGAGAGCAACATGTTTTTCCAGAG 39701

Qy   1002 CTATAGGGATGAAGCTCGTCGAGCTGCAGCTCGACGTAATCAAAGAAGCTGAAAATCGC 1061
          ||| |||||||||||||||||||||||  |||| ||||||||||||||||||||||||
Db  39700 CTTCAGGGATGAAGCTGAATCCGCGTCGGCCCGACGTAATCAGAGAGAGCTGAAAATCGC 39641

Qy   1062 TAGAGGAAAGAGAGGCCGATGTCGAGTAGATATGAGCCTGATTGATGAGGATAATCAAAACTGA 1121
```

```
Db  39640 CAGAGGAAAGAGAGCCATGTCAAGAAGATATGAGCTGATTGATGAGGAAATCGAGACTGA 39581
          |||||||||| |||||| |||||||  |||||||||||||| ||||||||| ||||||||
Qy   1122 GTATGAGCCAGAGTCATGGCGCCAAGGAGACGAAGCTACTGACACAATCCGACGAGGTTAC 1181
          ||||||||||||| ||||||||| |||  |||||||||||| ||||||||||||||||||
Db  39580 GTATGAGCAAGAGTCATGGCGCAGGGAGACGAAGCTGCTGACAAACCCGACGAGGTTAC 39521

Qy   1182 AGTGGAGGAGTATATCAGATTCTTTGAGATGAATGACTTCTCGGGGACGAGGTATCCCTG 1241
          ||||||||||||||||| ||||||||||  |||| |||||||| ||||||||||||| ||
Db  39520 AGTGGAGGAGTATATCAGGTTCTTTGAGGCTAATGACTTCTAGGGACGAGGTATCCGTG 39461

Qy   1242 ATATGAGACTTTAGCCCAGTTGGGGTTACTGGAGACGTGCAGCATCTGTTCGAGAAGTG 1301
          |||||||||||||||||||    ||||||||||||||| ||  |||||||||||||
Db  39460 CTATGAGACTTTAGCCCAGCCAGTGGGGTTACTGGAGACGTACAACATCTGTTCGAGAGGG 39401

Qy   1302 TCAT--------------CTGATAAGGAGGAGGACAATCGAGTT 1330
          ||||              |||| || |||||||||||| |||||
Db  39400 TCATATGGAGACACTCATGTCTTATCCCTTATCCGCTTATAGGAGGAGGACAACCGAGTT 39341

Qy   1331 TCTTTTCCACACTGCAAGTGGAAATGTATGAGGGACTCACAGAGACTTTGAGCTGGATACCAT 1390
          |||||||||| ||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db  39340 TCTTTTCCACGCTGCAAGTGGAGATGTATGAGGGACTCACAGAGACTTTGAGCTGGATACCAT 39281

Qy   1391 GGGGTTAGGCTTCTTGACGTTTGGTTTCCCTAGTGGAGTGATTCGTGATAACAGCGGTAC 1450
          |||||||||||||||||||||||||||||| ||||||||| ||||| | ||||||
Db  39280 GGGGTTAGGCTTCTTGACGTTCTCAGTCCACTATAGAGGGCCACTATAGGCAACAACAGCGGTAC 39221

Qy   1451 ATTGGAAGAGAACGTGTTTGGTTCCCAGTTGAAAAGGGAACCAACCCCAGGTTTGCAGGGGA 1510
          ||  ||||||| ||||||||||||||||||||||||||||||||||| | ||||||
Db  39220 GTTAGAAGAGAACGTGTTTGGTTCCCTAGTGGAAAGGGAACCAAACCCAGATTTGCACAGAGA 39161

Qy   1511 AGAGCTTAAGGATTGTGTGGGCTACTATTGGGGCAACAATCTACCGCTAAACTCGACCGGTC 1570
          |||||||||||||||||||||||||| ||| |||||||||||||||||||||||||||
Db  39160 GGAGCTTAAGGATTGTGTGGGCCACTATAGGGCAACAATCTACCGCTAAACTCGGCCGGTC 39101

Qy   1571 CAAGAGCAACCAACCAGTCCGAGTCTGTGATTCGCTACTTTCAGCGCCTCGGTTGCCAATGT 1630
          |||||||||||||||||||||||| ||||||||||||||||||||| |||||||| |||
Db  39100 TAAGAGCAACCAACCAGTCCGAGTCTGTGATCCGCTACTTTCAACGATCGGTTGCCAACGT 39041

Qy   1631 TTTTTACTCCAGGGAGTCTACAGGCACCGTGTCTAACACAGACATGAAGATGATAGATTC 1690
          |  || |||||||||||||||||||||||||||||| ||||||| ||||||||||||||
Db  39040 CTTTTACTCCAGGGAGTCTACAGGCACCGTGTCTAACAACAGAGATGATAGATTA 38981

Qy   1691 AGCGCTTATAGGGGATTCTCCGCGCCTTACAAAAGGAAAGATGTCCTGAGAGGATCTTAA 1750
          ||| |||| ||||||||||||||||||||||||||||||| || ||||||||||||
Db  38980 TGCGCTTATTGCGCGGATTCTCCGCCCGGACGAGGGGAAGAATGTCTTGAGAGGGATCTCAA 38921

Qy   1751 CGACTCACCACCAGTAATGCCTCTGTTGGATCATCTGTGTGGGTACATGAAGTGGGCGCT 1810
          ||||||||||||||||| |||||||||| ||||||||||||||||||||||||||||
Db  38920 CAACGCCACCAGTAATGCCTCTGTTGGATCATCTGTGTGGGTACATGAAGTGGGCCGTT 38861

Qy   1811 GACAAACGGCAAGAGAAGAGGTAAGGAGCACTATGCCAACATTCTTTCCCCTGCATTAACGCTAC 1870
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db  38860 GACGAACGGGAAGAAGAGGTGAGGAGCACTATGCCAACATTCTTCTCCCCTGCATTAACGCTAC 38801

Qy   1871 TCTGAAAGTTGTGGAGTTCCGCTCAAGGAGGTTAGCACCAGGAATGATGACTT 1930
          ||||||||||| |||||||||||||||||||||| |||  ||||||||||||||
Db  38800 TCTACAAGTTATGTGAGTTCTGCGCTCAAGGAGGTTCAGCAACCAGGATGATGGACTT 38741

Qy   1931 GGATCACTTGCGCCGATGGTTCTGAGTTTGACATGGTTGGCGACTTTGGCGCCGCTA 1990
          ||||||| |||||| |||||| |||||||||||||||||||||||||| |||||||
Db  38740 GGATCATTTGCGCCGATGGTTCTTGGAGTTTGACATGGTTGGCGACTTTGCATC 38681

Qy   1991 CAGGTTCGAGCATTCATCGATTAGAAATCGCCAACATTCTTTCCCCTGCATTTACGCTAC 2050
          |||||||||| |||||||||||||||||| || |||||||||||||||||||||||||||
Db  38680 CAGGTTCGAGCATTCATCGATCGAATCGCCAACATTCTTCTCCCCTGCATTCACGCTAC 38621

Qy   2051 TAGGATTCTGAGGGCAGGAACATTGACTTCAAGCCTTGCGCCTTCAAGAGGGATCAATGACTTCGA 2110
          ||||||||||||||||||||||||||||||||||||| ||||| ||||||| |||||||||||||
Db  38620 TAGGATTCTCGAGGGCAGCAACCATTGACTTCAAGCCTTGCGCCTTCAAGAGGGATCTTTATTTCGA 38561

Qy   2111 GGGCAGTCGCCAACTGAGGAGATTAGTAGTCACACCGAAGGAGCTACAATAGAAGATGTTGA 2170
          |||||||||| ||||| ||||||||  |||   |||| |||||||||||||||||| ||||
```

```
Db  38560 GGCTAGTCGCCAACTGAGGAGATCAGTTTACCGAAGGAGCTACAACAAAGATATTGA 38501
          ||||||||| |||||  ||||||||||| ||||| |||||||||||| |||||||||||
Qy   2171 TGAGACATATGATATAGATGAGGCGGAGTTGACACGAGCATTCAGTGAGCA 2230
          ||||| |||||||| ||||||| || |||||||| || |||||||||||||
Db  38500 TGAGACTGATGATAGATGATGAGGCTGAGTTTGACACGAGCATGTTTCAGTGAGCA 38441

Qy   2231 TATACCTCCAGCGACTGAATGACTTCTCGGGGACGAGGTATCCCTG 2290
          ||||||||||||||||||||||||||| ||||||||||||||| |||
Db  38440 TATACCCTCCTGCCGAGGGAAAGCAAGAGTTGAGCGGAAGCTCACAGGAACAACAGCAAGCT 38381

Qy   2291 GCAGAAGTGGTCAAGAACAGGATAAGTAGTTACTCGCCAAGTGCCTCAGGGCTATCAAGTT 2350
          |||||||||||||| |||||||||||||| ||||  ||||||||||||||||||||||||
Db  38380 GCAGAAGTGGTGCATGAACAGGACAAGTACTCGCCAAGTGCCTCAGGGCTATCAAGTT 38321

Qy   2351 TCTGAAGGACAAGATCAGCTGCTCCTTCCACTACAACTATTCCGCAATGACAGCTCCC 2410
          |||||||||| |||||||||  |||||||| |||||| ||||||||||| ||||||||
Db  38320 TATGAAGAACAAGCTCAGCTGCACCTCTTCTACTATAGCTATTCCGCAGGGCAATCTCCC 38261

Qy   2411 TCAGGACATGCCTTCGAGGAGATATGACGCGCCC------------GAGCC 2449
          ||||||||||||||||||||||||||||||||||            |||||
Db  38260 TCAGGACATGCCTTCGAGGAGATATGACGCGCCGCTCGGCCCGAGCCCAGCCATGAGCC 38201

Qy   2450 TAGAGAGCAGAAGATTCTGCATGTCCCTGCGAGGCATTCATCATTCGAGCCTTCGTGATC 2509
          |||||||||||||||||||||||||||| ||||||||||||||||| ||||||||||||
Db  38200 TAGAGAGCAGGAGATTCCGCATGTCCCTGCGAGGCATTCATCATTCGAGCCTTCGTAAATC 38141

Qy   2510 TAGGAAGAATAGGAACGACACTCACTGATCTAGCAGCAGCAGGACGACTTCTGCA 2569
          ||||||||| ||||| ||||||||||||||||||||  ||||  ||||||||||||
Db  38140 AGAGAGGAGGAGGAGGAACGACACCACTCGCTCGATCTCAGCAGCGGACGACTTCTGCA 38081

Qy   2570 GTCTCGTAGTTTACGCGACCGGCGTGCTGCCCGGTGCAATAGAAGAGAGGTCGAGTATCC 2629
          ||||||||||| |||  ||||||||||||| ||||||||||||||||||| |||||||||
Db  38080 GTCTCGTAGTTTACACGACCGTGCTGCTGCCCGGTGCAATAGCAGCGAGAGGAGGTCGAGTATCC 38021

Qy   2630 TCAGACGCGGTGCTGGCCCACAGAGCTGATGAGATCGAGTTCGGAGCTGA 2689
          |||||||||||||||||||||||| ||||||||| |||||||| ||||||
Db  38020 TCAGACGCGGTGCTAGCCCACATGATGATGAGTAGTGATGAGGTGTCGGAGCTGA 37961

Qy   2690 TACGGAACATGCGGCCTCGTCGTATGGCTGTCGTTCATCTATGCCGC 2749
          |  ||||||||||| ||||||||||||||| |||||||||||||||||
Db  37960 TACAGAACAAGGTGTTCATCTATGCCGGAGCAATCAGCAGCCATTGACGAGCA 37901

Qy   2750 ACTTCGTTCATTATTCGACTGAGGTAAGCCTCGAGTAATATATCATCTC 2809
          |||| |||||||||||||||||||||| ||||| |||||||||||||||
Db  37900 ACTACGTTCATTCTTCGAGTACTGAGGTAAGCCTCACTTACTTCATTGTATTTACCATCTA 37841

Qy   2810 TTGTGATTTGTT------CTTTATTTTGTTCTTCAGTGATTGCTCCTGAGTACTTCTCT 2864
          |||||||||||       ||||| ||| || |||||||||| ||||| ||||||||||
Db  37840 TTGTGATTTGTTTTCCTTTTCATTTTTTGTTCTTCAGTGATTGGTCCTGAGTACTTCTCT 37781

Qy   2865 TCCAAGTTTATTCACCAGTGGTTGTAAGGAGGTTAAGTTTGGGGAGGAGGGCTCCGGAAGT- 2923
          |||||||||||||||||| ||||||||| ||||||||||||||||| ||||||||||||||
Db  37780 TCCATGTTTATTCACCAGTGGTTGTAAGGAGGTTAAGTTTGGGGAGGAGGGCTCCGGAAGTG 37721

Qy   2924 -ATGTTGCATTGTATATATTTTTAAGTCTGCATTCATCTAAGGCATAGAAAAACCAAAAA 2982
          |||||||||||| ||||||| |||  ||||||||||||||||||||||||||| ||||||
Db  37720 TGTGTTGCATTGTGTATATTTTTGAGTCTGCATTCATCTAAGGCATAGAAAAACCAAAAA 37661

Qy   2983 AAAATT-AAAAATTTCAGAAAAATGATTTCAC-AAAAAAAGAGTGTTCATGTAGTTGCATT 3040
          ||||||  |||||| ||| ||||  |||||||  |||||||| ||||||||||||||||||
Db  37660 AAAATTGAAAAATTTTCGGAAAATGATTTTCCAAAAAAAAAGAGTGTTCATGTAGTTGCATC 37601

Qy   3041 ACATTTAGGATCAAGTCTAGAGTGTTGTTCATTGTTCATGCATAGGGGATAA 3100
          |||||||||||||||||||||||||||||| |||||||||||||| ||||||
Db  37600 ACATTTAGGATCAGTCTAGAGTGTTCATTGTTCATTTAGGATTGCATATGCGATGA 37541

Qy   3101 TGATGAGATAGCCTTGTAAGCA 3122
          |||||| ||||||||||||||||
Db  37540 TGATGAGATAGCCTTGTGAAGCA 37519
```

RESULT 8
T17A2/c

```
LOCUS           T17A2                  81902 bp    DNA     linear   PLN 10-AUG-1999
DEFINITION      Arabidopsis thaliana BAC T17A2.
ACCESSION       AF160183
VERSION         AF160183.1  GI:5106767
KEYWORDS        .
SOURCE          Arabidopsis thaliana (thale cress)
  ORGANISM      Arabidopsis thaliana
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
                rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE       1  (bases 1 to 81902)
  AUTHORS       Cordes,M., Woldmann,P. and Spalding,L.
  TITLE         The sequence of A. thaliana T17A2
  JOURNAL       Unpublished
REFERENCE       2  (bases 1 to 81902)
  AUTHORS       Washington University Genome Sequencing Center.
  TITLE         The A. thaliana Genome Sequencing Project
  JOURNAL       Unpublished
REFERENCE       3  (bases 1 to 81902)
  AUTHORS       Waterston,R.
  TITLE         Direct Submission
  JOURNAL       Submitted (17-JUN-1999) Department of Genetics, Washington
                University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE       4  (bases 1 to 81902)
  AUTHORS       Waterston,R.
  TITLE         Direct Submission
  JOURNAL       Submitted (10-AUG-1999) Department of Genetics, Washington
                University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT         Submitted by:
                Genome Sequencing Center
                Department of Genetics, Washington University,
                St. Louis, MO 63108, USA
                e-mail: rwilson@watson.wustl.edu

                MAPPING: Clones were assigned to the YAC map by hybridization by
                M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted
                by M. Marra, WashU, to pick the best candidates for sequencing.

                NOTICE: This sequence may not be the entire insert of this clone.
                It may be shorter because we only sequence overlapping sections
                once, or longer because we provide a small overlap between
                neighboring submissions.

                This sequence was finished as follows unless otherwise noted:
                all regions were double stranded or sequenced with an alternate
                chemistry; an attempt was made to resolve all sequencing problems,
                such as compressions and repeats; all regions were covered by
                sequence from more than one subclone

                        NEIGHBORING COSMID INFORMATION:

                Actual start of this clone is at base position 1 of T17A2; actual
                end is at 81902 of T17A2

                NOTES:

                Coding sequences below are predicted from computer analysis, using
                the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES             Location/Qualifiers
     source          1..81902
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="IV"
                     /map="unknown"
                     /clone="T17A2"
     gene            1601..3428
                     /gene="T17A2.2"
     CDS             join(1601..1741,1851..1982,2226..2386,2480..3428)
                     /gene="T17A2.2"
                     /note="contains similarity to Pfam family PF00155 -
                     Aminotransferases class-I; score=584.1, E=8.8e-172, N=1;
                     strong similarity to 1-aminocyclopropane-1-carboxylic acid
                     synthases"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAD48074.1"
                     /db_xref="GI:5724770"
                     /translation="MLSSKVVGDSHGQDSSYFLGWQEYEKNPFHESFNTSGIVQMGLA
                     ENQLSFDLIEKWLEEHPEVLGLKKNDESVFRQLALFQDYHGLPAFKDAMAKFMGKIRE
                     NKVKFDTNKMVLTAGSTSANETLMFCLANPGDAFLIPAPYYPGFDRDLKWRTGVEIVP
                     IHCVSSNGYKITEDALEDAYERALKHNLNVKGVLITNPSNPLGTSTTRREBDLLLTFT
                     STKKIHMVSDEIYSGTVFDSPEFTSVLEVAKDKNMGLDGKIHVVYSLSKDLGLPGFRV
                     GLIYSNNEKVVSAATKMSSFGLISSQTQHLLANLLSDERFTTNYLEENKKRLRERKDR
                     LVSGLKEAGISCLKSNAGLFCWVDLRHLLKSNTFEAEHSLWTKIVCEVGLNISPGSSC
                     HCDEPGWFRVCFANMSDQTMEVAMDRVKGFVDNNNGGKQKRTMWDTRRRSLINKWVSK
                     LSSVTCESER"
     repeat_region   complement(6823..6962)
                     /note="Athila_retroelement"
     repeat_region   complement(7214..7279)
                     /note="Athila_retroelement"
     repeat_region   complement(7237..7358)
                     /note="106B"
     repeat_region   complement(7299..7330)
                     /note="Athila_retroelement"
     repeat_region   complement(7361..7480)
                     /note="106B"
     repeat_region   complement(7470..7548)
                     /note="Athila_retroelement"
     repeat_region   complement(7473..7583)
                     /note="106B"
     repeat_region   complement(7581..7631)
                     /note="106B"
     repeat_region   complement(7585..7620)
                     /note="Athila_retroelement"
     repeat_region   complement(7631..7654)
                     /note="106B"
     repeat_region   complement(7692..7718)
                     /note="Athila_retroelement"
     repeat_region   complement(843..8722)
                     /note="Athila_retroelement"
     repeat_region   complement(8792..8828)
                     /note="Athila_retroelement"
     repeat_region   complement(8802..8828)
                     /note="H12_Athila_180_bp_repeat"
     repeat_region   complement(8826..8881)
                     /note="Athila_retroelement"
     repeat_region   complement(8826..8881)
                     /note="H12_Athila_180_bp_repeat"
     repeat_region   complement(8882..8913)
                     /note="Athila_retroelement"
     repeat_region   complement(8882..8913)
                     /note="H12_Athila_180_bp_repeat"
     repeat_region   complement(8919..9023)
                     /note="H12_Athila_180_bp_repeat"
     repeat_region   complement(8934..9023)
                     /note="Athila_retroelement"
     repeat_region   complement(9380..9446)
                     /note="Athila_retroelement"
     repeat_region   complement(9444..9600)
                     /note="H12_Athila_180_bp_repeat"
     repeat_region   complement(9444..9584)
                     /note="Athila_retroelement"
     repeat_region   complement(9603..9661)
                     /note="Athila_retroelement"
     repeat_region   complement(9603..9661)
                     /note="H12_Athila_180_bp_repeat"
     repeat_region   complement(9667..9751)
                     /note="Athila_retroelement"
     repeat_region   complement(9667..9750)
                     /note="H12_Athila_180_bp_repeat"
```

```
repeat_region    complement(9853. .10015)
                 /note="Athila retroelement"
repeat_region    complement(9853. .10015)
                 /note="H12 Athila 180 bp_repeat"
repeat_region    complement(10061. .10122)
                 /note="Athila retroelement"
repeat_region    complement(10061. .10122)
                 /note="H12 Athila 180 bp_repeat"
gene             complement(10148. .10122)
                 /gene="T17A2.3"
CDS              complement(join(10148. .10553,10663. .12148,13399. .13527,
                 13624. .15049,15162. .16001))
                 /gene="T17A2.3"
                 /note="contains similarity to retrotransposons; may be a
                 pseudogene"
                 /codon_start=1
                 /evidence=not_experimental
                 /protein_id="AAD48078.1"
                 /db_xref="GI:5724774"
                 /translation="MAEEQNQQNGPANIGAGDAPRDHRQRKGIAPPAIQNNNFEIKSG
                 LISMIQGNKFHGLPMEDPLDHLDEFDRLCNLTKINGVSEDGFKLRLFPFSLGDKAHIW
                 EKNLSHDSITTWDDYKKAGLFSKFFSNARTARLRNEIYGFSQKTGESFCRAWERFKGYT
                 NQCPHSFTKASLLSTLYRGVLPRIRMLLDTASNGNFQNKDVEEGWELVENLAQSDGN
                 YNEDCDRTVRGTADSSDDKHRKEIKALNDKLDRILLSQQKHVHFLVDDEQYQVDGEGN
                 QLEEVYLPQQKQGONKPFVLYNQGFVPKQQFQGNYQPPPPPGFVTQQNOCHAAPDAKM
                 KQMVKQLLOGQASSSMEIAKKLSELHHKLDLSYNDLNAKVEALNTKVRILEGQSASTF
                 SPKVTRLPGKSIQNPKEYATAHAIICHDRELPTRHVLDLITRDNDVQEGEASTOVEA
                 SVVEFNHSAGSRHLTQSTSEEKAAIIERMVKRFKPTPLPLRALPWTFRKAWMERYKSV
                 AAKQLDEIEAVMPLMEVLNLILDPHKVVRNLILERIKMYHDSDESDATPSRAADKRI
                 VQEKLEDPGSFTLPCSIGEFAFSDCLCDLDGASVNIMPLSMARRLEFIQYKPCDLTLIL
                 ADRSSRKHFGMLKDLPVMINGVEVPTDFVVLDMEVEHKDPLILGRPFLASVGAVIDVK
                 EGKIGLNLGKHIKLQFNINRTPQGSTEDGRTSGNDRVISREEYETERVKEFKKRYDKQ
                 DETIEKLAHTVEERRNYPLBEKEAYFBEKGIEYSAANFSRENAEYDDIREDYADLLY
                 SLIMSNYSGESSMDDYNVDEVESWSARPREOHVYQSFDEFERSTARHNQRRAEIAR
                 GKRAMSSRYELIDEDIETEYEPESWRKETKLLNKPDEVTVEEYIRFFEINDFWGTRYS
                 CYETLAQLRLLEDVQYLFBKCHLETLMSYPYPAYKEETIEFLSTLQVEMYEALTDFEL
                 DTMGLGFLTFLVDEQRYQLSIKKLEELFGPSGKTKPFPRDREELKDlWATIGNNLPL
                 NSARYKSNOIWSPVIRYFQQSVANVFYSKESTGTVSNTNMEMIDYALIGILRRTKGKN
                 VLRGDLNNAPPLMPLLIHICJVRKWALITNGKKKVRGALCVGGVVTPILEACGVSLKEP
                 GLAPRMDLDHLRRCEFLEPDMAGDFHRYRFEHSSIRIANILLPCIDATRILEGRNID
                 FKPALEDLYFKGSPPTDEISHTEGATTEDVDETDDIDEAEFDTSMYHFSEHIPPARES
                 KSLSEAHRNNKLQKWCKKQDKLLAKCLRAISAGLSLESRRFRMSRDIHHSSLVNLG
                 GRGGPHSLDLAAGADDFCSLVVTAVLAAAEGKRLSIIRAVLAATELMMSSTHLLEL
                 IQNKVVRPWPGSNHKQPLMSNYVHSSTEVSASLHHCIISSVVICFFILFL"
repeat_region    complement(10435. .10851)
                 /note="Athila retroelement"
repeat_region    complement(10435. .10851)
                 /note="H12 Athila 180 bp_repeat"
repeat_region    complement(11116. .11527)
                 /note="Athila retroelement"
repeat_region    complement(11366. .11833)
                 /note="H12 Athila 180 bp_repeat"
repeat_region    complement(13234. .13423)
                 /note="Athila_retroelement"

Query Match    62.0%;  Score 1937.2;  DB 8;  Length 81902;
Best Local Similarity  84.8%;  Pred. No. 0;
Matches 2289;  Conservative   0;  Mismatches 343;  Indels 66;  Gaps    8;
```

```
        12374 TCCTTAATAGCCTAGTCGTTTTAACCTCTCAAGCTCCATCCCTTCACTCGGCCGCGTCCAAA 12315
Db

Qy        723 TACACTCGACCAACAACTTCAAAGTAAAAAAAAGAAAAAAAAGTCAAAACCGACA 782
              ||||||| |||||||  |||||      |||||||||||||
        12314 ACCACTCGATCAACAATCCACAAACGAAAAAATAAAATAAATCAAAACCGGTTTTGGTA 12255
Db

Qy        783 GTTTCACTCAACCGGTTACTCGACCGGTAC---GCTGGTTGAGATTGTGTTTTTGGTTT 839
              |||||||||||||||||||||||||||||    |||||||||| ||| ||| |||||
        12254 T---ACTCGACCGGTTACTCGACCGGTTTTTTTGTGTTGGTTGGTGGTTTTGTTTTGGTTT 12199
Db

Qy        840 TGCTATTACTAACATATTAACGTTTATCTTTGAGTTCGTCTGTTTTAGGTTTCATCAT 899
              |||||||||||| ||||||||| |||||||||||||||||   |||||||||
        12198 TGCTGTTACTAACGTTATTATCGTTTATCTTTGAGTTTGTTGTTTTTCAGGTTCTCATCAT 12139
Db

Qy        900 GAGTAACTACAGTCGAAAATCCTCTATGGACCCTGATTATAATGTGGATGAAGCTAAGTC 959
              ||| ||||||||||||||||||||  |||||||||||| ||||  |||||| ||||||
        12138 GAGCAACTACAGTCGGAGATCTTCTATGGACCCTGATTACAAATGTGGATGAAGTTGAGTC 12079
Db

Qy        960 CTGGTCCACTAGACCGGAGTGAGCAACATGTTTACGAGCTAGAGAGCTATAGGGATGAATTTGA 1019
              |||||||||||||||||    ||||||||||||||||||||||||||||||||||
        12078 CTGGTCCGCTAGACC---AAGAGAGCAACATGTTTACCAGAGCTTCAGGGATGAATTCGA 12022
Db

Qy       1020 ACGCTCTGCAGCTCGACGTAATCAAAGAAGAGAGCTGAAATCGCTAGAGAAAGAGGGCGAT 1079
              |||||||||||||||||||||| ||||||||||||||||   |||||||||||||||||
        12021 ACGCTCTACAGCCCGACATAATCAGAAGAGAGCTGAAATCGCCAGAGAAAGAGGGCGAT 11962
Db

Qy       1080 GTCGAGTAGATGAGTGAGTGAGATATCAAAACTGAGTATGAGCCAGAGTCATG 1139
              |||||||||||||||||||||||||||| |||||||||||||||||||||||
        11961 GTCAAGTAGATATGAGTGAGTGATGAGGATATCGAAACTGAGTATGAGCCAGAGTCATG 11902
Db

Qy       1140 GCGCAAGGAGACGAAGCTACTGACAAATCCGACGAGGTTACAGTGGAGGAGTATATCAG 1199
              || ||||||||||||||| |||||||||||||||||||||||||||||||||||||||
        11901 GCGAAAGGAGGACGAAGCTGCTGAACAAACCTGACGAGGTTACAGTGGAGGAGTATATCAG 11842
Db

Qy       1200 AATTCTTTGAGATGAATGAACGAGAGTATCCCTGATCGTGCTATATGAGACTTTAGCCCA 1259
              ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
        11841 GTTCTTTGAGATAAATGATTTCGTGGGAACGCGGTATTCGTGCTATATGAGACTTTAGCCCA 11782
Db

Qy       1260 GTTGGGGTTACTGGACGTGCAGCATCGTTCGGAGAAGTGTCATCTG--------- 1308
              |   ||||||||||| |||||||||||||||||||||||||||||
        11781 GCTGAGGTTACTGGACGTGCAGTATCTGTTCGAGAAGTGTCACCTGGAGACACTCAT 11722
Db

Qy       1309 --------------ATAAGGAGGAGCAATCAGCTGAGTTTCTTCCACACTCGCAAGT 1348
                            |||||||||||||||||||||||||| |||||||||||||
        11721 GTCTTATCCTTATCCTGCCTATAAGGAGGAGCAATCGAGTTTCTTCCACGCTGCAGGT 11662
Db

Qy       1349 GGAAATGTATGGAGGACTCACAGACTTGAGCTGGATACCATGGGGTTAGGCTTCTTGAC 1408
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
        11661 GGAGATGTATGGAGGGCACTAACAGACTTGAGCTGGATACACCATGGGGTTAGGCTTCTTGAC 11602
Db

Qy       1409 GTTCTTAGTGGATGAAGAACAGCCGTACCAGATTTGAATCAAGAAATTGGAAGAACTGTTTGG 1468
              ||||||||||||||||||||||| ||| |||||||| ||| |||||||||||||||||
        11601 ATTCTTAGTGGATGAAGAACAGCGGTATCAGCTTTCGATCAAGAAGAACTCGGAAGAACTGTTTGG 11542
Db

Qy       1469 TTTCCCTAGTGGAAAGGGAACAACCCCAGTTTGACAGGGAAGAGAGCTTAAGGATTGTGTG 1528
              ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
        11541 TTTCCCTAGTGGAAAGGGAACCAACAAACCCAGATTTGACAGAGGAGAGCTTAAGGATTTTGTG 11482
Db

Qy       1529 GGCTACTATTGGGAACAATCTACCGTCAAACTCGACGCGGTCCAAGCAACCAAAATCCG 1588
              |||| |||||||||||||||||| ||||||||||||||||||||||||||||||||||
        11481 GGCCACTATAGGGAACAATCTACCCGGCGACTCACGCGGTCCAAGCAACCAAAATCTG 11422
Db

Qy       1589 GAGTCCTGTGATTCCTACTTTCAGCGCTCGGTTGCCAATGTTTTTTACTTCCAGGGAGTC 1648
              ||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
        11421 GAGTCCTGTGATCCGGATCCGCTACTTTCAGCAATTCGGTTGCCAACGTCTTTTACTTCCAAGGAGTC 11362
Db

Qy       1649 TACAGGCACCGTGTCTCAAACAGCAGCTCAACATGAGAGATGAGATTCAGCGCTTATAGGGATTCT 1708
              ||||||||||| ||||| |||||| |||||||||||||| ||||||||||| |||||||||||||
        11361 TACAGGCACCGTGTCTAACACGAACCATGGAGATGAGATGAGATTATGCGCTTATAGGGATTCT 11302
Db

Qy       1709 CCGGCCTTACAAAAGCAAAGAAATGCTTCTTGGACGAAGAGATCTTAACGACTCACCACCAGTAAT 1768
              ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
        11301 CCGTCGGACGAAGGGACGAAGAATGCTTCTTGGAGAGGCGGATCTCCAACAACGCACCACTAAT 11242
Db
```

```
Qy        483 CAACCAAACTCGACCTCCGTTCTCTCTGCCACTCATGATCGTCGATCTCTCCACCATC 542
              ||||  ||||||||||||  |||||  |||||||||||||| ||||||||||||||
        12553 CAACTCACTCGACCGGTTCCTCCTCTTCCGCCTTCACAGTCACTCGGCTCGGCCAAGCCA-C 12495
Db

Qy        543 TTCATCATCTCCCTTACTCGACACCGTGCGTCTCGCTCGACCATTGCCATTTAAAAGCT 602
              ||||||||||||||| |||||||||||||||||||||| ||||||||||||||  ||
        12494 ACCATCTCTCCCATCTCTTTGTCACTCGATCGGTCTACACACCATCAACAACTACTCGAC 12435
Db

Qy        603 CACTCGGATTGTCAAAGAGAAGAAGAGAGAGTCGAAGCTCAACACCACCGCCACTCGACCGGTTTCC 662
              ||||||||||||||||||  |||| |||||||||||||||| |||||| |||||||||||
        12434 CTCCTGCTCTTTTCTCAAGGAAGCAGCTCAACCTCCGCCGTTCAACTCCGATTCACTCGGCCACAGC 12375
Db

Qy        663 CTCTACACATTCAACACTCGACCACCGGTGCTACCATCTCCACATCTTGCCACCTCTTGTTCACCA 722
              |||||  ||||||||||||||||  |||     ||||||||||
        11301 CCGTCGGACGAAGGGACGAAGAATGCTTCTTGGAGAGGCGGATCTCCAACAACGCACCACTAAT 11242
Db
```

Thu Sep 9 11:17:23 2004

us-10-088-384a-27.rge

```
Qy    2827 TTTTGTTTTCAGTGATTGGATTGTTCCTGAGTACTCTTCCAAGTTTATTCACACAGTGG 2886
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10161 TTTTGTTTTGTGATTGGATTGTTCCTGAGTACTCTTCCAAGTTTATTCGCACAGTGG 10102

Qy    2887 ACTGTGTGATTTAAGTTTGGGGAGGGGGCTCCAGGAGTATG--TTGCATTGTATATATTT 2944
           ||||||||||||||||||||||| |||||||||||||||  ||||||||| |||||||||
Db   10101 ATTGTGTAATTTAAGTTTGGGGAGGGGGCTCCAGGAGTGTGTATTGCATTGTATATATTT 10042

Qy    2945 TAAGTCTGCATTCATCTAAGGCATAGAAAAAAATTCAGAAAATTAAAAATTTCAGAAAAT 3004
           ||||||||||||||||||||||||||||||||||  |||||||| ||||||||||||
Db   10041 TGAGTCTGCATTCATCTAAGGCATAGAAAAAAATCAAAAAATGAAAATTTCAGAAACT 9982

Qy    3005 GATTTCACAAAAAAGAGTGTTCATGTAGTTGCATTACATTTAGGATCAAGTCAGAGTG 3064
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9981 GATTTCACAAAAAAGAGTGTTCATGTAGTTGCATTACATTTAGGATCGAGTCTTGAGTG 9922

Qy    3065 TTTCATTTAGGATTGTTGCATATGCATAGGGGATAATGATGGATAGCCTTGTAAGCA 3122
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9921 TTTCATTTAGGATTGTTGCATATGCATAGGGGATGATGATGAGATAGCCCTTGTAAGCA 9864
```

RESULT 9
ATCHRIV21/c
LOCUS   ATCHRIV21   195837 bp   DNA   linear   PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 21.
ACCESSION   AL161509
VERSION   AL161509.2 GI:7267442
KEYWORDS   .
SOURCE   Arabidopsis thaliana (thale cress)
  ORGANISM Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 42621 to 43431)
  AUTHORS   Robben,J., Grymonprez,B., Volckaert,G., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   2 (bases 43282 to 43283)
  AUTHORS   Lamar,B., Stoneking,T., Stumpf,J., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   3 (bases 109648 to 109649)
  AUTHORS   Spiegel,L.A., Huang,E.N., Nascimento,L.U., de la Bastide,M.,
            Vil.D.M., Preston,R.R. Matero,A., Shah,R., O'Shaughnessy,A.,
            Rodriguez,M., Shekher,M., Schutz,K., See,L.H., Swaby,I.,
            Habermann,K., Dedhia,N.N., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   4 (bases 1 to 195837)
  AUTHORS   EU Arabidopsis sequencing project.
  TITLE   Direct Submission
  JOURNAL   Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie. Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT   Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATCHRIV20 at the 5' end and an
            overlap with ATCHRIV22 at the 3' end.
FEATURES   Location/Qualifiers
     source   1..195837
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     repeat_region   12352..12656
                     /note="function=near centromeric, similar to near
                     centromeric repeat mil67, GenBank accession number X94367"
     repeat_region   12651..12915
                     /note="function=near_centromeric, similar to near

```
Qy    1769 GCCTCTGTTGATCCATCTGTGTGGGTACATGAAGTGGCGCTGACAAACGGCAAGAGAGAA 1828
           ||||| ||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11241 GCCTCTCATTGATCCATCTGTGTGGATACAGGAAGTGGGCGTTGACGGAACGGGGAAGAGAGA 11182

Qy    1829 GGTAAGAGGAGCACTATGCGTGGGTGGCGTTGTGACGGCCAATTCTGAAAGTTTGTGAGT 1888
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   11181 GGTGAGAGAGCACTATGCGTGGGTGGCGTTGTGACGGCCAATTCTCGAAGCTTGTGAGT 11122

Qy    1889 TCCGCTCAAGGAGTAGGGTTAGCACCGAGAATGATGACTTGGATCACTTGGCCGATG 1948
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11121 ATCGCTCAAGGAACCAGGGTTAGCACCGAGGATGATGACTTGGATCATTTGCGCAGATG 11062

Qy    1949 TGAGTTCTCTGAGTTTGACATGGTTGGCGACTTTGCCGACTTTCACCGCTACAGGTTCGAGCATTCATC 2008
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11061 TGAGTTCTTGGAGTTTGACATGGCTGGCGACTTTCACCGCTATAGTTCGAGCATTCATC 11002

Qy    2009 GATTAGAATCGCCAACATTCTTTTCCCTGCATTTACGCTACTAGGATTCTCGAGGGCAG 2068
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10942 GATTAGAATTGCCAACATTCTTCTCCCCCTGCATTGATGCTACTAGGATTCTCGAGGGCAG 10942

Qy    2069 GAACATTGACTTCAAGCCTTGCGCTTGAAGATCTTTATTTCGAGGGCAGTCCGCCAACTGA 2128
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10941 GAACATTGACTTCAAGCCTTGAAGATCTTTATTTCAAGGGCAGTCCGCCAACTGA 10882

Qy    2129 GGAGATTAGTCACACCGAAGGAGCTACAATAGAAGATGTTGATGAGACATATATATAGA 2188
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10881 TGAGATCAGTCACACCGAAGGAGCTACAACAGAAGATGTTGATGAGACACGATATATAGA 10822

Qy    2189 TGAGGCGGAGTTTGACACAGCATGTATCATTCAGTGAGCATATACCTCCAGCGGAGGAA 2248
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10821 TGAGGCGGAGTTTGACACGAGTGTATCATTTTAGTGAGCACATACCTCCAGCGGAGGGA 10762

Qy    2249 AAGCAAGAGTTTGAGCGGAGCTCACAGGAACAACAGCAAGTGCGAGAAGTGGTGCAAGAA 2308
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10761 AAGCAAGAGTTTGAGCGGAGCTCACAGGAACAACAGCAAGTGCGAGAAGTGGTGCAAGAA 10702

Qy    2309 ACAGGATAAGTTACTCGCCAAGTGCCTCAGGGCTCATCAAGTTTCTGAAGGACAAGATCAG 2368
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10701 ACAGGATAAGTTACTCGCTAAGTGCCTCAGGGCTCATCAGGTTTCTGAAGGACAAGATCAG 10642

Qy    2369 CTGCTCCTCTTCCACTACAACTATTCCGCAATGACAGCTCCCCTCAGGACATGCCTTCGAG 2428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10641 CTGCTCCTCTTCCACTACAGCTATTCCGTAGGGACAGCTCCCCTCCGGACATGCCTTCGAG 10582

Qy    2429 ----------------------------GAGCCTAGAGACAGCAGAAGATTCT 2467
                                        ||||||||||||||||||||||||||
Db   10581 GAGATATGACGGCGCCCTGAGCCTAGAGAGCCTAGAGAGCAGCAGAGATTCC 10522

Qy    2468 GCATGTCCCTGCGAGGCATTCATCATTCGAGCCTCGTGAATCTAGGAAGAATAGGAGAAC 2527
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10521 GCATGTCCATGCGAGACATTCATCATTCGAGCCTCGTGAATCTGGGAGGAAGAGGAGGAGAC 10462

Qy    2528 GACACTCACTCGATCTAGCAGCAGCAGCAGCAGACGACTTCGTAGTTTACGCGA 2587
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10461 CACACTCACTCGATCTAGCAGCAGCAGCAGACGACTTCGTAGTTTACGCGA 10402

Qy    2588 CCGCGGTGTCGCCGCGCAATGAGGGTCGAGTATCCTCAGGACGCGTCGCGCCG 2647
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10401 CCGCGGTGTCGCCGCGCAGCAGAGGTTAAGTATCCTCGAGGAGCGCGTGCTGCGCCG 10342

Qy    2648 CCACAGAGCTGATGAGATCGACTGAGTACCCACCATGCTGGAGCTGATACGGGAACATGGCGGTTC 2707
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10341 CCACAGAGCTGATGATGTCGAGTTCGAGTACCCACCGTGGGAGCTGATACGAGCAAGGTGGTTC 10282

Qy    2708 GTCTATGGCTTGGGAGGCAATCACACGAGCAGCAATCACAGCAACTTCGTTCATTATTCGA 2767
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10281 GTCCATGGCTTGGGAGGCAATCACAGCAGCAATCACAGCAACTACGTTCATTCATTCTTCGA 10222

Qy    2768 CTGAGGTAAGCGCCTCACTTCACCTATTATTATTATCATCTCTTGTGATTTG-TTCTTTA 2826
           ||||||||||||||||||||||||||||||||||||||||| |||||||| |||||
Db   10221 CTGAGGTAAGCGCCTCACTTCACCTTCACCATTGTATTCATCCGTTGTGATTTGTTTCTTCA 10162
```

```
repeat_region   13035. .13165
                /note="function=near_centromeric, similar to near
                centromeric repeat mi167, GenBank accession number X94367"

repeat_region   13186. .13255
                /note="function=near_centromeric, similar to near
                centromeric repeat mi167, GenBank accession number X94367"

repeat_region   13376. .13506
                /note="function=near_centromeric, similar to near
                centromeric repeat mi167, GenBank accession number X94367"

gene            14391. .15248
                /gene="AT4g08030"

CDS             14391. .15248
                /gene="AT4g08030"
                /note="similarity to"
                /codon_start=1
                /product="putative Athila-like protein"
                /protein_id="CAB81140.1"
                /db_xref="GI:7267443"
                /db_xref="SPTREMBL:Q9ZQB2"
                /translation="MSTGRFSSSLCDLGSSINLMPKFVDERLGMTNYRPTRITLLFVD
                RSKRIPEGILEDVSVKVGNSLIPADFVVLDYDKEPKDPLILGRAFIATAGARIDVKKG
                RISLKVCDLEMEFGMDCSEFTTISSIASGTNTPPKTAQNPQTETHTMLPSAQLANES
                CRATMSIDTIPPVDRHPRDSQTEDSPHSLQPTPLAALYDKFSSISCPAFPSINSLDSS
                KSASQDNKPRVLRKLVSPVVLAGHEESIDRHHHECQSTPAAALTRSVYTPPWKARRQQ
                QSSTTPPSS"

exon            14391. .15248
                /gene="AT4g08030"
                /number=1

repeat_region   15252. .15354
                /note="function=centromeric_repeat, similar to 163A
                centromeric repeat, GenBank accession number X93608"

repeat_region   15376. .16133
                /note="function=centromeric_repeat, similar to 164A
                centromeric repeat, GenBank accession number X92080"

repeat_region   16235. .16337
                /note="function=unclassified, similar to a portion of the
                Athila retrotransposon"

repeat_region   16665. .17457
                /note="function=unclassified, similar to T24G3, GenBank
                accession number AC006192, similar to T15F16 GenBank
                accession number AF076275"

repeat_region   18166. .18236
                /note="function=unclassified, similar to a portion of the
                Athila retrotransposon"

repeat_region   20901. .21791
                /note="function=putative_LINE, similar to K21C13, GenBank
                accession number AB010693"

repeat_region   22551. .23015
                /note="function=unclassified, similar to T25C13, GenBank
                accession number AF080121"

repeat_region   27565. .27723
                /note="function=centromeric_repeat_fragment, similar to a
                portion of 278A centromeric repeat, GenBank accession
                number X92081"

repeat_region   28408. .28953
                /note="function=near_centromeric, similar to near
                centromeric repeat mi167, GenBank accession number X94367"

repeat_region   33513. .33832
                /note="function=unclassified, similar to T29E15, GenBank
                accession number AC005170, similar to MUL8, GenBank
                accession number AB009054"

repeat_region   34525. .34797
                /note="function=unclassified, similar to K21C13, GenBank
                accession number AB010693"

repeat_region   35559. .36770
                /note="function=unclassified, repeat consists of three
                segments, similar to T27D20, GenBank accession number
                AF076274, similar to T2L5, GenBank accession number
                AF096371"

repeat_region   37837. .38095
                /note="function=putative_LINE, similar to T3F12, GenBank
                accession number AC002983"

repeat_region   39806. .40044
                /note="function=unclassified, similar to MRO11, GenBank
                accession number AB005244, similar to MXA21, GenBank
                accession number AB005247, similar to MOP10, GenBank
                accession number AB005241"

repeat_region   40169. .40342
                /note="function=unclassified, similar to T17A5, GenBank
                accession number AF024504"

gene            44883. .46710
                /gene="AT4g08040"

CDS             join(44883. .45023,45133. .45264,45508. .45668,45762. .46710)
                /gene="AT4g08040"
                /note="contains similarity to Pfam family PF00155 -
                Aminotransferases class-I, score=584.1, E=8.8e-172, N=1,
                strong similarity to 1-aminocyclopropane-1-carboxylic acid
                synthases
                strong similarity to
                Contains Aminotransferases class-I pyridoxal-phosphate
                attachment site AA264-277
                contains EST gb:Z33841"
                /codon_start=1
                /protein_id="CAB81141.1"
                /db_xref="GI:7267444"
                /db_xref="GOA:Q9S9U6"
                /db_xref="SPTREMBL:Q9S9U6"
                /translation="MLSSKVVGDSHGQDSSYFLGWQEYEKNPFHESFNTSGIVQMGLA
                ENQLSFDLIEKWLEHPEVLGLKKNDESVFRQLALFQDYHGLPAFKDAMAKFMGKIRE
                NKVKFDTNKMVLTAGTSANETLMFCLANPGDAFLIPAPYYPGFDRDLKWRTGVEIVP
                IHCVSSNGYKITEDALEDAYERALKHNLNVKGVLITNPSNPLGTSTTREELDLLITFT
                STKKIHMVSDEIYSGTVFDSPEFTSVLEVAKDKNMGLDGKIHVVYSLSKDLGLPGFRV
                GLIYSNNEKVVSAATKMSSFGLISSQTQHLLANLLSDERFTTNYLEENKKRLRERKDR
                LVSGLKEAGISCLKSNAGLFCWVDLRHLLKSNTFEAEHSLWTKIVCEVGVLNISPGSSC
                HCDEPGWFRVCFANMSDQTMEVAMDRVKGFVDNNNGGKQKRTMWDTFRRRSLINKWVSK
                LSSVTCESER"

exon            44883. .45023
                /gene="AT4g08040"
                /number=1

intron          45024. .45132
                /gene="AT4g08040"
                /number=1

exon            45133. .45264
                /gene="AT4g08040"
                /number=2

intron          45265. .45507
                /gene="AT4g08040"
                /number=2

exon            45508. .45668
                /gene="AT4g08040"
                /number=3

intron          45669. .45761
                /gene="AT4g08040"
                /number=3

exon            45762. .46710
                /gene="AT4g08040"
                /number=4

repeat_unit     50105. .50244
                /note="Athila_retroelement"

repeat_unit     50496. .50561
                /note="Athila_retroelement"

repeat_unit     50519. .50640
                /note="106B"

repeat_unit     50581. .50612
                /note="106B"

repeat_unit     50643. .50762
                /note="Athila_retroelement"

repeat_unit     50752. .50830
                /note="106B"

repeat_unit     50755. .50865
                /note="Athila_retroelement"

repeat_unit     50863. .50913
                /note="106B"

repeat_unit     50867. .50902
                /note="Athila_retroelement"
```

```
repeat_unit      50913..50936
                 /note="106B"
repeat_unit      50974..51000
                 /note="Athila_retroelement"

Query Match      62.0%;  Score 1937.2;  DB 8;  Length 195837;
Best Local Similarity  84.8%;  Pred. No. 0;
Matches 2289;  Conservative  0;  Mismatches 343;  Indels 66;  Gaps 8;

Qy    483  CAACCAAACTCGACCTCGTCTCTTGCCACTCGATAGTCACTCGATCTCTCCTCCACCATC   542
           |||||||  |||  ||||||||||| |||||||| ||||||||  |||||||||||||||
Db  55835  CAACTTCACTCGACCGCTCCTCTCCTGCCTTTCACAGTCACTCGACTTCGCCAAGCCA-C   55777

Qy    543  TTCATCATCTCCCTTACTCGACCACCGCGTCGCGTCTCGCTCGCTCCACCATTGCCATTT   602
           ||||||| ||||| ||||||||||||||||||||||| |||| ||||||||||||||||
Db  55776  ACCCATCTCCCATCTCTTTGTCACTCGACTCGATCGCGTCTACACACCATCAACAACTCGAC   55717

Qy    603  CACTCGATTGTCAAAGAGAAGAGAGAAGAGTGAAGCTCAACCTCGGCCACTCGGCCGTTCC   662
           ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db  55716  CTCCTGCTCTTTTCAAAGAGAAGAGAGAAGAGTGAAGCTCAACCTCGCCGTTCCCACACTA   55657

Qy    663  CTCTACACATTCAACACTCGACCACGTGTCACCATCTCCAAGTCCCATCCCTTCACTCGGC   722
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55656  TCCCTTAATAGCCTAGTCGTTTTAACCTCCAAGTCCCATCCCTTCACTCGGCCGGTCCAAA   55597

Qy    723  TACACTCGACCACCAACTTCAAAGTAAAAAAAAGAAGAAGTCAAAACCGACA   782
           |||||||||||||||||||||||||||||||||
Db  55596  ACCACTCGATCACAACAATCCACAACAAAAAAAATAAATCAAAACCGGTTTGGTA   55537

Qy    783  GGTAAGAGAGCACTATGCGTGGCGTTGTGACGCCAATTCGAAGTTGTGAGT   839
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55536  T---ACTCGACCGGTTTACTCGACCGGTAC--GCTGGTTTAGATTGTGTTTTGGTTT   55481

Qy    840  TCCGCTCAAGGAAGTAGGGTTAGCACCGAGAATGATGGATCTTATTCGAGGGCAGTCCGATG   899
           ||||||||||||||||||||||||||||||||
Db  55536  TGCTATTACTAACATATTACGTTTATCTTGAGTTCGTCGTTTGTTTGTTTCATCAT   55421

Qy    900  TGAGTTCTCTGAGTTTGACATGTTCGCGACTTTCACCGTACAGGTTCGAGCATTCATC   959
           |||||||||||||||||||||||||||||||||||||
Db  55420  GAGTAACTACAGTCGAAATCCTCATGGACCCTGATTATAATGTGGATGAAGCTAAGTC   55361

Qy   1019  GATTAGAATCGCCAACATTCTTTCTTCCCCGCATTTAGCTACTAGGATTCTCGAGGGCAG   2068
           |||||||||||||||||||||||||||||||||
Db  55360  CTGGTCCACTAGACCGAGTGAGAGCAACATGTTACGAGCGTATAGGAATGAATTTGA   55304

Qy   1079  GAACATTGACTTCAAGCCTCGCCTGACTTGAAGATCTTTATTCAAGGGCAGTCCGCCAACTGA   2128
           ||||||||||||||||||||||||||||||||||
Db  55303  ACGCTTCGCAGCTCGACGTAATCAAAGAAGAGCTGAAATCGCTAGAGAGAAAGAGGGCGAT   55244

Qy   1139  GGAGATTAGTCACACGAAGGAGCTACAATAGAAGATGTTGATGAGACAATATGACGATATAGA   2188
           ||||||||||||||||||||||||||||
Db  55243  ACGCTCTACACGCCCCGACATATATGTTGATGAGGATATCAAAAACTGAGTATGAGCCAGAGTCATG   55184

Qy   1199  TGAGGCGGAGTTTGACACGAGTGTATCATTCAGTGACGAGCATATACCTTCCAGCGAGGAA   2248
           ||||||||||||||||||||||||||||||||||
Db  55124  GCGCAAGGAGAGACGAAGCGAAGCTACTGAACAAATCCGAACTCGAGGAGTGAGAGTATATCAG   55124

Qy   1259  AAGCAAGAGTTTGAGCGAAGCTCAGGGTATCCCTGATATGAGACTTTAGCCCA   2308
           ||||||||||||||||||||||||||||||
Db  55064  GCGAAAGGAGAGACGAAGCGAAGCTACTGAACAAATCCGAGGAGTGAGAGTATATCAG   55064

Qy   1308  ACAGGAGAAGTTACTCGCCAAGTGCCTCAAGTTCTGAAGGACAAGAGATCAG   2368
           ||||||||||||||||||||||||||||||
Db  53984  ATTCTTTGAGATGAATGCATAAATGATTTCGTGGGAACGCGCGGTATCCCGTATTCGTGCTAATGCTATTAGCCCA   53984

Qy   1309  -----------AATAAGGAGGAGACAATCGAGTTTCTTTCCACTGCAAGT   2428
           ||||||||||||||||||||||||||||||
Db  53924  GTTCTTTTGAGATAAATGATTTTGAGCTGCTATGCTATGCTAATGCTATTAGCCCA   53924

Qy   1348  CTGCTCCTCTTCCACTACAACTAGTGAGTTTCTTTCCACCGCTGCAGGT   2467
           ||||||||||||||||||||||||||||||
Db  55004  GTTGGGGTTACTCGGAGGACGTGCAGCATCTGTTCGACGAAGTGCATCTG-------CAG   55004

Qy   1408  GGAAATGTATGAGGGACTTTGAGCTGGATCACCATGGGGTTAGGCTTCTTGAC   53804
           ||||||||||||||||||||||||||||||
Db  55063  GTTGGGGTTACTCGGAGGACGTGCAGCATCTGTTCGACGAAGTGCATCTG-------CAG   55063

Qy   2467  GAGATATGACGCGCCC---------------GAGCCTAGAGACGCAGAAGATTCT   2467
           |||||||||||||||
Db  53863  GAGATATGACGCGCCGCCGAGGACGCCGAGCCTAGAGACGGAGAGAGAATAGGAGAAC   53804

Qy   2468  GCATGTCCCTGCCGAGGCATTCATCATTCGAGCCTCGTGAATCTAGGAAGAATAGGAGAAC   2527
Db  54943  GGAGATGTATGAGGGCACTAACAGACTTGAGCTGGATGAATACCATGGGGTTAGGCTTCTTGAC   54884
```

Right column (alignment continuation):

```
Qy   1409  GTTCCTTAGTGGATGAACAGCGGTACCAGATTTAGAGTCATCAAGAAATTGGAAGAACTGTTTGG   1468
           ||||||||||||||||||||||||||||||
Db  54883  ATTCTTAGTGGATGAACAGCGGGTATCAGCTTCGATCAGCAAGAAGTTGGAAGAACTGTTTGG   54824

Qy   1469  TTTCCCTAGTGGAAAGGGAACCACCCACCCAGGTTTGACACGGGAAGAGCTTAAGGATTTGTG   1528
           ||||||||||||||||||||||||||||||
Db  54823  TTTCCCTAGTGGAAAGGGAACCACCCAACCAGATTTGACAGAGAGAGCTTAAGGATTTGTG   54764

Qy   1529  GGCTACTATTGGGAACAATCTACCGCTAAACTCGACACGCGGTCCAAGAGCAACCAAATCCG   1588
           ||||||||||||||||||||||||||||||
Db  54763  GGCCACTATAGGCACACAATCTACCGCCGGTCCAATGTTTTTACTCCAGGGAGTC   54704

Qy   1589  GAGTCCTGTGATTGCTACTTTCAGCGCGTCGGTTGCCAATGTTTTTACTCCAGGGAGTC   1648
           ||||||||||||||||||||||||||||||
Db  54703  GAGTCCTGTGATCCGTCACTTTCAGCAATCGGTTGCCAACCTCTTTTACTCCAAGGAGTC   54644

Qy   1649  TACAGGCACCGTGTCTAACGAACATGAAGATGATGATTCAGCGCTTATAGGGATTCT   1708
           ||||||||||||||||||||||||||||||
Db  54643  TACAGGCACCGTGTCTAACGAACATGAAGATGATGATTATGCGCTTATAGGGATTCT   54584

Qy   1709  CCGCTCGTTACAAAGGAAAGAAATGTCCTGAGAGAGAGATCTTAACGACTCACCCACCAGTAAT   1768
           ||||||||||||||||||||||||||||||
Db  54583  CCGTCCGACGAAGGGAAGAGAATGTCTTGAGAGAGGATCCTCAACACGTCCAACCACCACTAAT   54524

Qy   1769  GCCTCTGTTGATCCATCTGTGTGGGTACATGAAGTCGGGCTGACAAACGGCAAGGAAGAGAA   1828
           ||||||||||||||||||||||||||||||
Db  54523  GCCTCTATTGATCCATCCGTGTGTGGGTGGCGTTGTGACGCGGTTGACGGGGAAGGAAGAGAA   54464

Qy   1829  GGTAAGAGAGCACTATGCGTGGCGTTGTGACGCCAATTCTCGAAGTTGTGTGAGT   1888
           ||||||||||||||||||||||||||||||
Db  54463  GGTGAGAGAGCACTATGCGTGGCGTTGTGACGCGTTGTGACTTCTCGAAGCTTGTGTGAGT   54404

Qy   1889  TCCGCTCAAGGAAGTAGGGTTAGCACCGAGAATGATGGATCTTATTTCAAGGGCAGTCCGCCAACTGA   1948
           ||||||||||||||||||||||||||||||
Db  54403  ATCGGCTCAAGGAACCAGGGTTAGCACCGAGGATGATGGATCTTATTTCAAGGGCAGTCCGCCAACTGA   54344

Qy   1949  TGAGTTCTCTGAGTTTGACATGGTTCGCGACTTTCACCGTACAGGTTCGAGCATTCATC   2008
           ||||||||||||||||||||||||||||||
Db  54343  TGAGTTCTTGGAGTTTGACATGGTTGACATGGTTCACCGTATAGGTTTCGAGCATTCATC   54284

Qy   2009  GATTAGAATCGCCAACATTCTTTCTTCTCCCCGCATTTAGCTACTAGGATTCTCGAGGGCAG   2068
           ||||||||||||||||||||||||||||||
Db  54283  GATTAGAATTGCCAACATTGTTCTTCTTCTCCCCTGAAGATCTTTATTCAAGGGCAGTCCGCCAACTGA   54224

Qy   2069  GAACATTGACTTCAAGCCTCGCCTGACTTGAAGATCTTTATTCAAGGGCAGTCCGCCAACTGA   2128
           ||||||||||||||||||||||||||||||
Db  54223  GAACATTGACTTGACTTGACTTGAAGATCTTTATTCAAGGGCAGTCCGCCAACTGA   54164

Qy   2129  GGAGATTAGTGCACACGAAGGAGCTACAATAGAAGATGTTGATGAGACAATATGAGACGATATAGA   2188
           ||||||||||||||||||||||||||||||
Db  54163  TGAGATCAGTCACACGAAGGAGCTACAACAGATGTTGATGAGACGATATGAGACGATATAGA   54104

Qy   2189  TGAGGCGGAGTTTGACACGAGTGTATCATTCAGTGACGAGCATATCCTTCCAGCGAGGAGGAA   2248
           ||||||||||||||||||||||||||||||
Db  54103  TGAGGCGGAGTTTGACACGAGTGTATCATTTTGTGAGCACGAGTATGCATCCAGCGAGGAGGA   54044

Qy   2249  AAGCAAGAGTTTGAGGCGAAGCTCAGGTGGTGCAAGACAAGATCAG   2308
           ||||||||||||||||||||||||||||||
Db  54043  AAGCAAGAGTTTGAGGCGAAGCTCAACAACAGCTCGCAGAAGTGGTGCAAGACAAGATCAG   53984

Qy   2309  ACAGGAGAAGTTACTCGCCAAGTGCCTCAAGTTCTGAAGGACAAGATCAG   2368
           ||||||||||||||||||||||||||||||
Db  53983  ACAGGATAAGTTACTCGCTCAAGTGCCTCAAGTTTCAGTTTCTGAAGGACAAGATCAG   53924

Qy   2369  CTGCTCCTCTTCCACTACAACTATTCCGCACTCCACTGCAAGT   2428
           ||||||||||||||||||||||||||||||
Db  53923  CTGCTCCTCTTCCACTACAACTATTCCGGGTTAGGCTTCTTGAC   53864

Qy   2429  GAGATATGACGCGCCC--------------GAGCCTAGAGACGCAGAAGATTCT   2467
           |||||||||||||||
Db  53863  GAGATATGACGCGCCGCCGAGGACGCCGAGCCTAGAGACGGAGAGAGAATAGGAGAAC   53804

Qy   2468  GCATGTCCCTGCCGAGGCATTCATCATTCGAGCCTCGTGAATCTAGGAAGAATAGGAGAAC   2527
```

```
Db  53803 GCATGTCCATGCCGAGACATTCATCATTCGAGCCTCGTGTAAATCTTGGGACGGAAGAGGAGGAGAC 53744
          ||||||||| |||||| ||||  ||| ||| ||  |||| ||| |||||||||| |||| |||
Qy   2528 GACACTCACTCGATCTAGCAGCAGCAGACGACACTTCTGCAGTCTCGTAGTTTACGCGA 2587
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53743 CACACTCACTCGATCTAGCAGCAGCAGACGACACTTCTGCAGTCTCGTAGTTTACGCGA 53684

Qy   2588 CCGCGGTGCTGCGGCCGCAATAGAAGAAGAGAAGAGAGTCGAGTATCCTCAGAGCGGTGCTGGCCG 2647
          ||||||||||||||||||| | |||||||||||||||||  ||||||||||| ||||||||||||||
Db  53683 CCGCGGTGCTGCGGCCGCAGCAGAGCAGCGAAGGAAGAGAGTTAAGTATCCTGAGAGCGGTGCTGGCCG 53624

Qy   2648 CCACAGAGCTGATGAGATCGAGTACCCACATGCTGGAGCTGATACGGAACATGCGCGGTTC 2707
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53623 CCACAGAGCTGATGATGTCGAGTACCCACCTGCTGGAGCTGATACAGAACAAGGTGGTTC 53564

Qy   2708 GTCTATGGCTTGGGACGCAATCACAGGCGAGCCATTGACTACCAACTTCGTTCATTATTCGA 2767
          |||| ||||| |||| || ||||||||| |||||||||||||||||||| ||| |||| |||
Db  53563 GTCCATGGCCTCGGGAGCCAATCACAAGCAGCAGCCATTGATGAGCAACTACGTTCATTCTTCGA 53504

Qy   2768 CTGAGGTAAGCGCCCTACTTCACCATTATATATATCATCTTGTGATTTG-TTCTTTA 2826
          |||||||||||||||||||||||||||||||||||||||||||||| |||   ||||
Db  53503 CTGAGGTAAGCGCCCTACTTCACCATTCATCATCCGTTGTGATTTGTTGTTCTTTCTTCA 53444

Qy   2827 TTTGTTTCAGTGATTGGATTTGTCCTGAGTACTCTCTTCCAAGTTTATTCACACAGTGG 2886
          |||||||| |||||| ||||||||| | |||||||||||||||||| ||||||||||||
Db  53443 TTTTGTTTTTTGTGATTGGATTTGTCCTGAGTACTCTCTTCCAAGTTTATTCCCACAGTGG 53384

Qy   2887 ACTGTGTGATTAAGTTTTGGGGGAGGGCTCAGGAAGTATG--TTGCATTGTATATATTT 2944
          ||  ||||||||||||| |||||||||||||||||||||||  |||||||||||| |||||
Db  53383 ATTGTGTAATTTAAGTTTTAAGTTTGGGGGAGGGCTCAGGAAGTGTATTGCATTGTATATATTTT 53324

Qy   2945 TAAGTCTGCATTCATCTAAGGCATAGAAAAACCAAAAAAAAAAAATTAAAAAATTTCGAGAAAT 3004
          |||||||||||||||| |||||||||  ||||||||||||||||||   |||||  |||||||
Db  53323 TGAGTCTGCATTCATCTAAGGCATAGAAAAACAAAAAAAAAAAAATGAAAAAATTTCAGAAAACT 53264

Qy   3005 GATTTCACAAAAAAGAGTGTTCATGTAGTTGCATTACATTGAGGATCAAGTCTAGAGTG 3064
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53263 GATTTCACAAAAAAGAGTGTTCATGTAGTTGCATTACATTTAGGATCGAGTCTTGAGTG 53204

Qy   3065 TTTCATTTAGGATTGTTGCATATGCATAGGGGATAATGATGAGAAGCCTGTGAAGCA 3122
          ||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db  53203 TTTCATTTAGGATTGTTGCATATGCATAGGGGATGATGAGAGATAGCCTTGTGAAGCA 53146
```

misc_feature　　/organism="Arabidopsis thaliana"
　　　　　　　　/mol_type="genomic DNA"
　　　　　　　　/cultivar="Columbia"
　　　　　　　　/db_xref="taxon:3702"
　　　　　　　　/chromosome="4"
　　　　　　　　/map="centromere"
　　　　　　　　/clone="F15N16"
　　　　　　　　1..1000
　　　　　　　　/note="overlap to BAC F13J5, please refer for analysis and annotation"

misc_feature　　1000..5529
　　　　　　　　/note="similarity to fragments of Athila retrotransposons"

LTR
repeat_region　5530..7082
　　　　　　　　7083..20826
　　　　　　　　/note="similarity to 180 bp centromeric repeat"

repeat_region　20827..24935
　　　　　　　　/note="similarty to 160 bp repeat"

LTR　　　　　　24936..25193
LTR　　　　　　25194..27048
misc_feature　27049..34345
　　　　　　　　/note="similarity to fragments of Athila retrotransposons"

LTR　　　　　　34346..35940
misc_feature　35941..41422
　　　　　　　　/note="similarity to fragments of Athila retrotransposons"

LTR　　　　　　41423..43041
LTR　　　　　　43042..44457
LTR　　　　　　44458..45957
misc_feature　45958..53796
　　　　　　　　/note="similarity to fragments of Athila retrotransposons"

LTR
repeat_region　53797..55873
　　　　　　　　55874..58433
　　　　　　　　/note="similarty to 160 bp repeat"

LTR　　　　　　58434..58806
misc_feature　58807..62369
　　　　　　　　/note="similarity to fragments of Athila retrotransposons"

LTR
repeat_region　62370..62737
　　　　　　　　62738..78607
　　　　　　　　/note="similarty to 160 bp repeat"

CDS　　　　　　complement(79991..81205)
　　　　　　　　/codon_start=1
　　　　　　　　/evidence=not_experimental
　　　　　　　　/product="extensin-like protein"
　　　　　　　　/protein_id="BAB83614.1"
　　　　　　　　/db_xref="GI:18149208"
　　　　　　　　/translation="MMNHLLALFLVGTIALPDRGTQQLTNDGLTGARREILSFLDQR
LAVVYPVIQRFKSLITLDPYNVTKTWIGSDICSYRGFHCNDPHNKTAVTVASIDPNG
FQLSAPSIREGFIDQFADLALFHVNSNNFGGTVPSKIVNLRYLYELDISNNRFTGQFPT
AVVQMSGLTFIDIRRNSFGGSIPPQILGQNLEVLFINDNGFTASLPEIPGDGTTHILF
LTLANNKFNGPLPRSILRSMSTLITEVLFLNNDFTGCIPHIGFLTGASVIDIGNKLT
GPLPLSLMCLEKVEQLNFAGNLLFGAVPEAVCMLLRDNLVNLSLSDNYFTHVGPWCRG
LLEKGVLDVGNNCIPFFPGQRSMQECAEFFVKPKKYYCPHMWFHSFFSCRYSHISSSS
SSAFPMVAPSP"

CDS　　　　　　complement(83214..83666)
　　　　　　　　/codon_start=1
　　　　　　　　/evidence=not_experimental
　　　　　　　　/product="AP2 domain containing protein"
　　　　　　　　/protein_id="BAB83615.1"
　　　　　　　　/db_xref="GI:18149209"
　　　　　　　　/translation="MVIQYKRKQEFPMVKEGMVMTEKPKRNLISSNEKRYKGIRMRKW
GKWVAEIRRPNKRSRIWLGSYKTAVAAARAYDTAVFYLRGPSARLNFPEEVFKDGNGG
EGLGGDMSPTLIRKKAAEVGARVDAELRLENRMVENLDMNKLPEAYGL"

misc_feature　93163..95514
　　　　　　　　/note="similarity to fragments of En/Spm transposon"

ORIGIN

Query Match　　　　　　　　61.7%;　Score 1925.4;　DB 8;　Length 123460;
Best Local Similarity　89.2%;　Pred. No. 0;
Matches 2203;　Conservative　0;　Mismatches 201;　Indels 67;　Gaps　9;

```
Qy    716 TTCACCATACACTCGACCAACAACTCTCAAGTAAAAAAAAAAGAAAAAAAAAGTCAA 775
          ||| |||| || ||||| ||| | ||| || |||||||||||  |||||||||  |||
Db  50062 TTCCAATAATCACTCGATCAACTACCACCCATAAAACGAAAAAAAAAAATAAAAAAAATTCA 50121
```

```
Qy       776 ACCGACAGTT-----CACTCAACGGTTACTCGACCGGTACGC---TGGTTTAGATTGT 828
                ||||         |||||||||||||||||||||||       ||||||||||||
Db     50122 AAACCGGCTTGGTATACTCGACCGGTTACTCGACCGGTGTTTTGCTTTGTGTGGTTTGT 50181

Qy       829 GTTTTGGTTTTGCTATTACTAACATATTAACGTTATCTTTGAGTTTCGTCTGTTTTA 888
                ||||||| || ||||||||||||||||  ||||||||||||||||||||||||||
Db     50182 GTTTTTGGTTTTGTTGGAACTAACGTATTAACGTTATTTTTTGAGTTTTACCATTTCA 50241

Qy       889 GGTTTCATCATGAGTAACTACAGTGAAAAATCCTCTATGGACCCTGATTATAATGTGAT 948
                ||||||||||||||||||||||  |||||||||||||||||||||||||||||||||
Db     50242 AGTTTCGTAATGAGTAACTACAGTGGAGAATCCTCTATGGACCCTGATTACAATGTGGAT 50301

Qy       949 GAAGCTAAGTCCTGGTCCACTAGACCGGATGAGCAACAATGTTACGAGAGCTATAGG 1008
                ||||||||||||||||||||||        |||||||||||||||||||||||
Db     50302 GAGCAAGCAACATGTTACCAGAGCTTCAGG 50358

Qy      1009 GATGAATTTGAACGCTCTGCAGCTGCAGCCCGACGTAATCAAAGAGAGCTGAAATCGTCAGAGGA 1068
                ||||||||||||||| |||||| ||||||||||||||||||||||||||||||||||
Db     50359 GATGAATTCGACACGCTCTGCAGCCCGACGTAATCAGAGAGAGCTGAAAATCGTCAGAGGA 50418

Qy      1069 AAGAGGGCGATGTCAGTCGAGTATATGAGCTGATTGATTGAGTTATATACAAAACTGAGTATGAG 1128
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     50419 AAGAGGGCGATGTCAAGTAGAGTATATGAGCTGATTGATTGAGTTATGAGAATCGAAAACTGAGTATGAG 50478

Qy      1129 CCAGAGTCATGGCCAAGGAGACGAAGCTACTGAACAAATCCGACGAGGTTACAGTGGAG 1188
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     50479 CCAGAGTCATGGCCAAGGAGACGAAGCTCTGAACAAACCCGACGAGGTTACAGTGGAG 50538

Qy      1189 GAGTATATCAGATTCTTTGAGATGAATGAGATGAATGAGCGGTATCCCTGATATGAG 1248
                |||||||||||||||||||||||||||||||| |||| ||||||||||||||||
Db     50539 GAGTATATCAGGTTCTTTGAGATGAATGAGATGAATGAGCGGTATCCCTGCTATGAG 50598

Qy      1249 ACTTTAGCCCAGTGTGGGGTTACTGAGGAGTGTCATCTG 1308
                |||||||||||||||||||||||||||||||||||||||
Db     50599 ACTTTAGCCCAGTGTGGGGTTACTGAGGAGTGTCATCTG 50658

Qy      1309 ------------ATAGGAGGAGACAATCGAGTTTCTTTCC 1337
                            |||||||||||||||||||||||||||
Db     50659 GAGAGACACTCATGTCTTACCCTTACGCTGCTTATAAGGAGGAGACAATCGAGTTTCTTTCC 50718

Qy      1338 ACACTGTCAAGTGCAAGTGATGAGGGACTCACAGACTTTGAGCTTGAGTACCATGGGGTTA 1397
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     50719 ACGCTGTCAAGTGCAAGTGATGGAGATGGAGAGGCATTCAGCGTGCTGGATACCATGGGGTTA 50778

Qy      1398 GGCTTCTTGACGTTTCAGTGGTAACAGCGGTACCGGTATTAGATCAAGAAATTGGAA 1457
                ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     50779 GGCTTCTTGACGTTTCAGTGGGTTACTGATGAACAGCAGCGGTATCTTTGATCAAGAAGTTGGAA 50838

Qy      1458 GAACTGTTTGGTTCCCTAGTGCAAAGGGAACCAACCCCAGGTTTGACAGGGAAGCTT 1517
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     50839 GAAGAGTGTTTGGTTCCCTAGTGGAAAAGGGAACCAAACCCAGGTTTGACAGAGAGAGCTT 50898

Qy      1518 AAGGATTGTGTGGGACTACTATTGGGAACAATCTACCGGTCAAACTCACCGGCGTCCAAGAGC 1577
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     50899 AAGGATTGTGTGGGGACTACTATTGGGGAACAATCTACCGGTCAAACTCACCGGCGTCCAAGAGC 50958

Qy      1578 AACCAAATCCGGAATCAGCTTTCAGCGGCTCGGTTGCCAATGTTTTTTAC 1637
                |||||||||||||||||||||||||||||||||||||||||||||||||
Db     50959 AACCAAATCCGGAATCAGCTTTCAGCGGCTCGGTTGCCAATGTTTTTTAC 51018

Qy      1638 TCCAGGAGTCTACAGGACACCGTGTCTAACAGGAAGGAATGTCCTGAGAATGAAGATGTAGATC 1697
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     51019 TCCAGGAGTCTACAGGAGCATCGTGTCTAACAGGAACATGGAATGTCCTGAGAATGAGATAGATC 51078

Qy      1698 ATAGGGATTCTCCGCCTTACAAGAAGGAAAATGTCTTGAGGACGAAGATCTTAACGACTCA 1757
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     51079 ACAGGGATTCTCCCGGACGAACGGCGATCTCAACAACCCA 51138

Qy      1758 CCACCAGTAATGCCTCGTCGTTGATCATCGTGTCGGTACATGAAGTGGCGCTGACAAAC 1817
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     51139 CCACCAGTAATGCCTCGTCGTTGATCATCGTGTCGATCAGGAGTGGCGTTGACGAAC 51198

Qy      1818 GGCAAGAAGAAGGTAAGAGCAGCACTATCGCTGGCGTGGCCGTTGTGACGCAATTCTGAAA 1877
```

```
Db     51199 GGAAAGAAGAAGGTGAGAGGAGCACTATGCGTGGCGTGGTGCCGTTGTGACGCGAATGATGACTTGGATCAC 51258

Qy      1878 GTTTGTGGAGTTCCGCTCAAGGAAGTAGGGTTAGCACCGGAAGAATGCGTGGGTGGCCTTGAGACTTGGATCAT 1937
                ||||||||||||||||||| || |||||||||||||||||||||||||||||||||||
Db     51259 GCTTGTGGAGTACCGCTCAAGGAACCAGGGTTAGCACTGGATTGGAATGAACTGGATCAT 51318

Qy      1938 TTGGCCGATGTGAGTTCTCTGAGTTGACATGGTTGGCGACTTGACGGCTACCGGCTACAGGTTC 1997
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     51319 TTACGCATATGTGAGTTCTCTGAGTTTGACATGGCTGGCGACTTGACGGCTACAGGTTC 51378

Qy      1998 GAGCATTCATCGATTAGAATCGCCAACATTCTTTCCCCTGCATTACGCTACTACTAGGATT 2057
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     51379 GAGCAATCATCGATTAGAATCGCCAACATTCTTCCCCCTGCATTGACGCTACTACTAGGATT 51438

Qy      2058 CTCGAGGGCAGGAACATTGACTTCAAG-CCTGCGCTTGAGACTCTTTATTCGAGGGCAG 2116
                ||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db     51439 CTCGAGGGCAGGAACATTGACTTCAAGTGCGCTTGAGAGATCTTTATTTCGAGGGCAG 51498

Qy      2117 TCCGCCAACTGAGGAGATTAGTCACCGACGAGGAGCTACAATAGAGAAGATGTTGATGAGAC 2176
                |||||||||||||||| || |||||||||||||||||||||||||||||||||||||||
Db     51499 TCCGCCAACTAAGGAGGATCAGTCACACAGAGGAGCTAGAACAGAGGAGCATGTTGATGAGAC 51558

Qy      2177 ATATGATATAGATGAGGCGGAGTTGACACCGAGTTGACACCGAGTTTGACGAGTCATTGAGCGATATACC 2236
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     51559 GGATGAAATAGATGAGGCGGAGTTTGACACCGAGTTGACACCGAGTGTATGTATCAGTGAGCATATACC 51618

Qy      2237 TCCAGGCGAGAAAAGCAAGAGTTTGAGCGAAGCAACAACAGCGAAGCTGCAGAA 2296
                |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     51619 TCCTGCGAGGAAAGCAAGAGTTTGACGAAGCTGCAAGTGCCTCAGGCGTATCAAGTGCGAGAA 51678

Qy      2297 GTGGTGCAAGAACACAGGATATAGTTACTCGCGACGCGGTATCATCGTGTCGAAGTTTCTGAA 2356
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     51679 GTGGTGCAAGAACACAGGATAAGTTACTCGCGACGCGGTATCATCGTGTCGAAGTTTCTGA 51738

Qy      2357 GGACAAGATCAGCTGCTTCCTCTTCCACTACAACTATTCCGCAATGATCAGCTCCCTCAGGA 2416
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     51739 GGACAAGATCAGCTGCTCCTCTTCCACTACAACATTCGTGATAGGGACAGCTCCCTCATGA 51798

Qy      2417 CATGCCTAGAGA 2455
                ||||||
Db     51799 CATGCCTTGAGGAGGATATGACGCGCCC 51858

Qy      2456 GCAGAAGAATTCTGCAGTCATGGTGCTCGTATGCTGTTGGAGCAAGTCATCATCATCGAGAAATTGGAA 2515
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     51859 GCAGGAGATTCCGGCATGGTGCTCGTATGCTGTTGGAGCAAGTCATCATCATCGAGCTGTGGAG 51918

Qy      2516 GAATAGGAGAACCACACTCACTCGATCTCAGCGGTTTGACAGGGAAGCTT 2575
                |||||||||||||||||||||||||||||||||||||||||||||||||
Db     51919 GAAGAGGAGAACCACACTCACTCGATCTCAGCGAGCAGCAGCCACTTCTGCAGTCTCAGTCTAG 51978

Qy      2576 TAGTTTACGGCACCGGCTGCTGCCGGCAAGGAGAGTGAGTCGGAGTCTCCAGAGC 2635
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     51979 TAGTTTACTGCAGCTGCTGCCGGCAAGGAGATCGGAGTCGGAGTCTCCAGAG 52038

Qy      2636 CGGTGCTGCCGCCACAGAGCTGATGAGTCGGAGTCGGAGTCTCCAGTCGATGCGGA 2695
                ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     52039 CGGTGTTGGCGGTTCGTCTATGCTCTATGCCTGTGGGAGCAATCGACTGACTCACAGAGCTGATGATACGGA 52098

Qy      2696 ACATGGGCGTTCGTCTATGCTCTATGCCTGTGGGAGCAATCGACAGCAGCAACTACGG 2755
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     52099 ACAAGGCCGGTTCGTCTATGCTCTATGCCTGTGGGAGCAATCGACAGCAGCAACTACG 52158

Qy      2756 TTTCATTATTCGACTGAGGTAAGCGCCTCACTTCACCATTATTACCATTGTATTATATATCTCTTCAAGTTGA 2815
                |||||||||||||||||||||||||| ||||||||||||||||||  || ||| |||||||||||||
Db     52159 TTTCATTCTTCGACTGGGGTTAAGCGCCTCAATTTACGTTGATTTGATTTGTGTTTATATATCTCTTTCCAAGTTGA 52218

Qy      2816 ATAGGGATTCTCCGCCGACGAAGGCGAAGAATGTCTTCAGTTGATTTGCGTGCTACTGGGCGCTGACAAAC 2874
                || | ||| ||| ||| | |||| |||||||| ||| ||| |||| ||||  ||| |||||
Db     52219 TTTGTTTCTTCATTTGTCGTTCCTTGATTTGCCTGGAGTCGTGTTTCCTCAGTGATACTCTCTTCCAAGTTTA 52278

Qy      2875 TTCACACAGTGGACTGTGTGATTAAGTTGCAT 2932
                || | ||| |||||| |||||  ||||||| ||
Db          TTCCACAGTGGACTGTGTGATTTAAGTTGTGCAT
```

```
Db  52279  TTCGCACAGTGGACTGTGTGATTTAAGTTTGGGGAGGGCTCAGGAAGTGTGTGTTGCAT  52338
           |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2933  TGTATATATTTTTAAGTCTGCATTCATTCTAAGGCATAGAAAAACCAAAAAAATTAAAA  2992
           |||  |||||||||||||||||||||||||||||||||||||||||||||  ||||
Db  52339  TGTGTATATTTTTGAGTCTGCATTCATTCTAAGGCATAGAAAAACCAAAAAAATTGAAAA  52398
           |||  |||||||||||||||||||||||||||||||||||||||||||||  ||||
Qy   2993  ATTTCAGAAAATGATTTCAC-AAAAAAGAGTGTTCATGTAGTTGCATTACATTTAGGAT  3051
           ||||||||||||||||||||  |||||||||||||||||||||||||||||||||||||
Db  52399  ATTTCAAAAAATGATTTCACAAAAAAAAGAGTGTTCATGTAGTTGTATTACATTTAGGAT  52458
           ||||| |||||||||||||  ||||||||||||||||||||||||||||||||||||||
Qy   3052  CAAGTCTAGAGTGTTCATTTAGGATTGTTGCATATCGCATAGGGATAATGATGAGATAG  3111
           ||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db  52459  CGATTCCACAGTGTTTTATTTAGGATTGTTGCATATACATAGGGATAATGATGAGATAG  52518
           |||||||||||||||||||  |||||||||
Qy   3112  CCTTGTAAGCA  3122
           |||||||||||
Db  52519  CCTTGTAAGCA  52529
```

RESULT 11
AC084241/c
LOCUS   AC084241   14879 bp   DNA   linear   PLN 19-JAN-2001
DEFINITION Arabidopsis thaliana chromosome 1 BAC F16M11 genomic sequence,
            complete sequence.
ACCESSION   AC084241
VERSION   AC084241.1 GI:10864177
KEYWORDS   HTG.
SOURCE   Arabidopsis thaliana (thale cress)
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 14879)
AUTHORS   Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J.,
            Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
            Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
TITLE   Arabidopsis thaliana chromosome 1 BAC F16M11 genomic sequence
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 14879)
AUTHORS   Town,C.D. and Kaul,S.
TITLE   Direct Submission
JOURNAL   Submitted (18-OCT-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
REFERENCE   3 (bases 1 to 14879)
AUTHORS   Town,C.D. and Kaul,S.
TITLE   Direct Submission
JOURNAL   Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
            Address all correspondence to:at@tigr.org
COMMENT   BAC clone F16M11 is from Arabidopsis thaliana chromosome 1
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of several methods: Gene
            prediction programs including Genscan+ (Chris Burge,
            http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky,
            http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant
            of GlimmerM, see Mihaela Pertea,
            http://www.tigr.org/softlab/glimmerm_htm/glimmerm.html, and
            GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
            mpertea@tigr.org), searches of the complete sequence against a
            peptide database and the plant EST database at TIGR
            (http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
            indicate the level of evidence for their annotation. Genes with
            similarity to other proteins are named after the database hits.
            Genes without significant peptide similarity but with EST
            similarity are named as unknown. proteins. Genes without protein
            or EST similarity, that are predicted by more than two gene
            prediction programs over most of their length are annotated as
            hypothetical proteins. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
            Simple repeats are identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html).

FEATURES   Location/Qualifiers
     source   1..14879
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="I"
                     /map="?"
                     /clone="F16M11"
     repeat_region   complement(1..1054)
                     /rpt_family="Rf:gi|7330786|gb|AF204874.1|AF204874
                     Arabidopsis thaliana conserved centromere sequence CCS1"
     misc_feature   1
                     /note="(TR) 60155 nt beyond this point were not included
                     in the submitted sequence due to an overlap with another
                     BAC (F19C17)"
     repeat_region   65..5003
                     /rpt_family="Rf:ATHILA|ATHILA A.thaliana DNA for Athila
                     retroelement.#Rf:ATR0033|X81801 Athila retroelement"
     repeat_region   complement(149..1157)
                     /rpt_family="Rf:gi|7330787|gb|AF204875.1|AF204875
                     Arabidopsis thaliana conserved centromere sequence CCS2"
     repeat_region   complement(149..417)
                     /rpt_family="Rf:centromeric repeat from T7M24 72424 to
                     72761"
     repeat_region   complement(1271..1403)
                     /rpt_family="Rf:gi|7330786|gb|AF204874.1|AF204874
                     Arabidopsis thaliana conserved centromere sequence CCS1"
     repeat_region   complement(1465..1677)
                     /rpt_family="Rf:gi|7330787|gb|AF204875.1|AF204875
                     Arabidopsis thaliana conserved centromere sequence CCS2"
     repeat_region   1805..2183
                     /rpt_family="Rf:ATGAGPOL1_I|ATGAGPOL1 I a consensus.
                     similar to del-like retrotransposon from T32N15,
                     GB|AC002534."
     repeat_region   1805..2180
                     /rpt_family="Rf:ATGP2I|ATGP2I Internal region of ATGP2
                     gypsy-like LTR retrotransposon."
     repeat_region   1805..2180
                     /rpt_family="Rf:TY3G|TY3G Ty3/gypsy LTR retrotransposon (a
                     consensus)."
     gene   1807..3349
                     /gene="F16M11.1"
                     /note="pseudogene, gypsy/Ty3-element polyprotein
                     (fragment)"
                     /pseudo
     repeat_region   complement(1832..2170)
                     /rpt_family="Rf:TEKAYI|TEKAYI Internal part of
                     gypsy/Ty3-type retrotransposon Tekay."
     repeat_region   1837..2000
                     /rpt_family="Rf:ATGP2I|ATGP2I Internal region of ATGP2
                     gypsy-like LTR retrotransposon."
     repeat_region   2819..3171
                     /rpt_family="Rf:ATGP2I|ATGP2I Internal region of ATGP2
                     gypsy-like LTR retrotransposon."
     repeat_region   2828..3261
                     /rpt_family="Rf:ATGAGPOL1_I|ATGAGPOL1 I a consensus.
                     similar to del-like retrotransposon from T32N15,
                     GB|AC002534."
     repeat_region   complement(2843..3267)
                     /rpt_family="Rf:ATR0059|AC002534 T32N15 del-like
                     retrotransposon"
     repeat_region   2843..3261
                     /rpt_family="Rf:TY3G|TY3G Ty3/gypsy LTR retrotransposon (a
                     consensus)."
     repeat_region   3349..5004
                     /rpt_family="Rf:ATHILA|ATHILA A.thaliana DNA for Athila
                     retroelement.#Rf:ATR0033|X81801 Athila retroelement"
     repeat_region   3349..5003
                     /rpt_family="Rf:Athila LTR for Athila retroelement"
     repeat_region   3349..3780
                     /rpt_family="Rf:gi|895701|emb|X89195.1|ATSATDNA1
                     A.thaliana DNA of a 180bp satellite
```

```
                    /note="(TF) 13602 nt beyond this point were not included
                    in the submitted sequence due to an overlap with another

Query Match        61.2%;   Score 1911.6;   DB 8;   Length 14879;
Best Local Similarity 90.1%;   Pred. No. 0;
Matches 2147;   Conservative 0; Mismatches 174; Indels 61; Gaps 7;

Qy    796 GGTTTACTCGACCGGTACGCTGGTTAGATTGTGTTTTGGTTTGCTATTATTAACATA 855
          ||||||||||||||||||| |||||||||||||||| ||||||||||||||||||||
Db  12663 GGTATACTCGACCGATTTACTCTGTTGGTTTGTGTTTTGGTTTAGTTGGGACTAACGTA 12604

Qy    856 TTAACGTTTATCTTGAGTTTCGTCTGTTTTTAGGTTCATCATGAGTAACTACAGTGGA 915
          |||||||||||||||||||| ||||||||||||| ||||||| |||||||||||||
Db  12603 TTAACGTTTATCTTGAGTTTCATTTGTTTTCAGGTTTCATTATGAGTAACTACAATGGA 12544

Qy    916 AAATCCTCTATGGACCCTGATTATAATGTGGATAAGCTAAGTCCTGGTCCACTAGACCG 975
          ||||||||||||||| |||||||| ||||||||||||||||||||||| |||||||||
Db  12543 GAATCCTCAATGGATCCTGATTATAATGTGGATAAAGCTAAGTCCTGGTCTTGGT-CACTAGACCG 12485

Qy    976 GAGTGAGAGCAACATGTTTACGAGCAGCTATAGGGATGAAATTGAACGCTCTGCAGCTCGA 1035
          |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db  12484 GAGAGAGAGCAACATGTTTACGAGCAGCTTCAAGGATGAAATTGAACGCTCCGCAGCTCGA 12425

Qy   1036 CGTAATCAAAGAAGAGCTGAAATCGCTAGAGAAGAGAGGGCGATGTCGAGTAGATATGAG 1095
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  12424 CGTAATCAGAGAAGAGCTGAAATCGCTATAGGAAAGAGGGTGATGTCGAGTAGATATGAG 12365

Qy   1096 CTGATTGATGAGGATATCAAAACTGAGTATGAGCCCAGAGTCATGGGCGCAAGGAGACGAAG 1155
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db  12364 CTGATTGATGAGGATATCGAAACTGAGAATCGAGAGTCATGGGCCAGAGTCATGGGCGCCGAAG 12305

Qy   1156 CTACTGAACAAATCGACGAGGTTACAGTGAGGAGTATATCGATGAAT 1215
          ||||||||||||||||||||||| |||||||||||||||||||||
Db  12304 CTGCTGAACAAACCCGACGAGGTTACAGTGAGGAGTATGTCAGATTCTTGAGATGAAC 12245

Qy   1216 GACTTCTGGGGAACGAGAGGTATCCCTGATATGAGACTTTGGGGTTTACTCGGAG 1275
          ||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db  12244 GACTTCTGGGGAACGAGAGGTATCCCTACTATGAGACTTTAGTCGAGCTGGAGCTGGGGTTACTCGGAG 12185

Qy   1276 GACGTGCAGCATCTGTTCGAGAAGTGTCATCTG------- 1308
          |||||||||||||||||||||||||||||||||
Db  12184 GACGTGCAGCATCAATTCGAGAAGTGTCATCTGGAGACACTCATGTCTACCCTTATGCT 12125

Qy   1309 ---ATAAGGAGGACAAATCGAGTTTCTTTCCACCTGCAAGTGGAAATGTATGAGGGA 1364
             ||||||||||||||||||||||||
Db  12124 GCCTATAAGGAGGACAAATCGAGTTCCTTT----ACTGCAAGTGGACGATGTATGAGGGT 12069

Qy   1365 CTCACAGACTTTGAGCTGGATACCATGGGGTTAGGCTTCTGACGTTCTTAGTGGATGAA 1424
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  12068 CTCACGGACTTTGAGCTGGATACCATGGGGTTAGGGTTCTGACGTTCTTCAGTGGATGAA 12009

Qy   1425 CAGCGGGTACCAGATTTAGATCAAGAAATTGGAGAAGAACTGTTGTTCCCTAGTGGAAAG 1484
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  12008 CAGCGGGTATCAGCTTTCGATCAAGTAGTTGGAGAAGAACTGTTTGTTTGTTTCACTAGTGGAAAG 11949

Qy   1485 GGAACCAACCCAGGTTTGACACGGAAGAGCTTAAGGATTGGTTGGCTACTATTGGGAAC 1544
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  11948 GGAACCAAACCCAGGTTTGACAAGGAAGAGCTTAAGGATTGGTTGGGCTACTATTGGGAAC 11889

Qy   1545 AATCTACCGCTAAACTCGACGCGGTCCAAGAGCAACCAAATCCGGAGTCCTGATTCGC 1604
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db  11888 AAT-TATCGCTAAACTCGGCACGGTCCAAAGCACCAACCAAATCCGGAGTCCTGTGTTGATCCGC 11830

Qy   1605 TACTTTCAGCGCTCGGTTGCCAATGTTTTTACTCCAGGGAGTCTACAGGCACCGTGTCT 1664
          ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db  11829 TACTTTCAGTGCTCGGTTGCCAACGTTTTCTACTCGGAGGGAGTCCACAGGCACCGTGTCT 11770

Qy   1665 AACACAGACATGAAGATAGATAGATTCAGCGCTTATAGGAGGACTTCCGCCTTACAAAAGGA 1724
          ||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db  11769 AACACAGACATGGAGATGAGATAGATTTCGGCGTTACAGGGAGGACTTCCGCCTTACAAAAGGGA 11710

Qy   1725 AAGAATGTCCTGAGGAGGATCTTAACGACTCACCACCACTAATGCCTCTTGTTGATCCAT 1784
          ||||||||||
Db  14879
```

repeat_region   junction#Rf:ATR0034|X89195 clone H12 of Athila 180-bp
                 repeat"
                 3349..3500
                 /rpt_family="Rf.gi|895702|emb|X89196.1|ATSATDNA2
                 A.thaliana DNA of a 180bp satellite
                 junction#Rf:ATR0035|X89196 5-end of clone H2 of Athila
                 180-bp repeat"

repeat_region   4600..5003
                 /rpt_family="Rf.gi|895704|emb|X89198.1|ATSATDNA4
                 A.thaliana DNA of a 180bp satellite
                 junction#Rf:ATR0037|X89198 clone H33 of Athila 180-bp
                 repeat"

repeat_region   4898..5003
                 /rpt_family="Rf.gi|895703|emb|X89197.1|ATSATDNA3
                 A.thaliana DNA of a 180bp satellite
                 junction#Rf:ATR0036|X89197 3-end of clone H2 of Athila
                 180-bp repeat"

repeat_region   complement(5103..5314)
                 /rpt_family="Rf:centromeric repeat from T7M24 72424 to
                 72761"

repeat_region   complement(5127..5324)
                 /rpt_family="Rf.gi|7330787|gb|AF204875.1|AF204875
                 Arabidopsis thaliana conserved centromere sequence CCS2"

gene   complement(6605..7236)
                 /gene="F16M11.2"
                 /note="pseudogene, gypsy/Ty3-element polyprotein
                 (fragment)"
                 /pseudo

repeat_region   complement(6650..6850)
                 /rpt_family="Rf:ATGP5I|ATGP5I Internal region of ATGP5
                 gypsy-like LTR retrotransposon."

repeat_region   complement(7372..7671)
                 /rpt_family="Rf:centromeric repeat from T7M24 69183 to
                 70294"

repeat_region   complement(7372..7464)
                 /rpt_family="Rf:ATHILA|ATHILA A.thaliana DNA for Athila
                 retroelement.#Rf:ATR0033|X81801 Athila retroelement"

repeat_region   complement(7372..7464)
                 /rpt_family="Rf:Athila LTR for Athila retroelement"

repeat_region   7672..8993
                 /rpt_family="Rf:centromeric repeat from T7M24 17342 to
                 19092 1751nt includes within it 106B centromeric repeat"

repeat_region   complement(7768..8165)
                 /rpt_family="Rf:106B centromeric repeat X93611, 106B is a
                 diverged copy of the Athila retroelement, 300
                 copies#Rf:gi|1100798|emb|X93611.1|AT106B A.thaliana
                 centromeric repeat region (clone 106B)#Rf:ATR0044|X93611
                 106B"

repeat_region   complement(8166..8992)
                 /rpt_family="Rf:Arabidopsis thaliana retrotransposon
                 romani At 5, reverse transcriptase-like sequence GB
                 AF073829"

repeat_region   complement(8949..9443)
                 /rpt_family="Rf:ATHILA|ATHILA A.thaliana DNA for Athila
                 retroelement.#Rf:ATR0033|X81801 Athila retroelement"

repeat_region   complement(10275..10439)
                 /rpt_family="Rf:ATHILA|ATHILA A.thaliana DNA for Athila
                 retroelement.#Rf:ATR0033|X81801 Athila retroelement"

repeat_region   complement(10277..10439)
                 /rpt_family="Rf.gi|895701|emb|X89195.1|ATSATDNA1
                 A.thaliana DNA of a 180bp satellite
                 junction#Rf:ATR0034|X89195 clone H12 of Athila 180-bp
                 repeat"

repeat_region   complement(13749..13891)
                 /rpt_family="Rf:ATHILA|ATHILA A.thaliana DNA for Athila
                 retroelement.#Rf:ATR0033|X81801 Athila retroelement"

repeat_region   complement(13917..14041)
                 /rpt_family="Rf:ATHILA|ATHILA A.thaliana DNA for Athila
                 retroelement.#Rf:ATR0033|X81801 Athila retroelement"

repeat_region   complement(14233..14871)
                 /rpt_family="Rf:ATHILA|ATHILA A.thaliana DNA for Athila
                 retroelement.#Rf:ATR0033|X81801 Athila retroelement"

misc_feature   14879

```
Db   11709 AAGAAATGTCTTGAGAGGGGATCTCAACAACGACCACCAGTAATGCCTCGTTGATCCAC 11650

Qy    1785 CTGTGTGGGTACATGAAGTGGGGGCCTGACAAACGGCAAGAAGAAGGTAAGAG-AGCACT 1843
           |||||| ||| |||||||||||||||| |||||||||||||||||||||| |||||
Db   11649 CTGTGTGATACAGGAGGTGGGGGCCGTTGATAAACGGCAAGAAGAAGGTGAGAGCAGCACT 11590

Qy    1844 ATGCGTGGGTGGCGTTGTGACGGCCAATTCTGAAAGTTTGTGGAGTTCCGCTCAAGGAAGT 1903
           ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   11589 ATGCGTGGGTGGCGTTGTGATGCGAATTTGGAGAGCCTTGTGGAGTACCGCTCAAGGAGCC 11530

Qy    1904 AGGGTTAGCACCGAGAATGATGATGACTTGGATCACTTGCGCCGATGTGAGTTCTCTGAGTT 1963
           |||||||||||||||||| |||||||||||||||||||||||||||| |||||||||
Db   11529 AGGGTTAGCACCGAGAATGATGATGACTTGGATCACTTGCGCCGCTGCTGATGTGAGTTCTCTGAGTT 11470

Qy    1964 TGACATGGTTGGCGACTTTCACCGCTACACAGGTTCGAGCATTCATCGATTAGAATCGCCAA 2023
           ||||||||||||||||| ||||||||||||||||||||||| ||||||||||||||||
Db   11469 TGACATGGTTGGCGACTTTCACCGCTACACAGATTCGAGCATTCATCGATTCGAATCGCCAA 11410

Qy    2024 CATTCTTTCCCCTGCATTTAGCTACTAGGATTCTCGAGGGCCAGGAACATTGACTTCAA 2083
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   11409 CATTCTTCTCCCCTGCATTTAGCTACTAGGATTCTCGAGCAAAGGAACATTGACTTCAA 11350

Qy    2084 GCCTGCGCTTGAAGATCTTTATTTCGAGGGCCAGTCCCCAACTGAGGGAGTTAGTCACAC 2143
           ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   11349 GCCTGCGCTTGAAGATCTTTATTTCGAGGGCCAGTCCGCAGACTGAGGGAGATTAGTCACAC 11290

Qy    2144 CGAAGGAGCTACAATAGAGATGTTGATGAGACATATGAGATGGCCGGAGTTTGA 2203
           |||||||||| |||||||||||| |||||||||||||||||||||||||||||
Db   11289 CGAAGGACCTACAAAAGAAGAGATGTTGATGAGACAGATGATGGCCGTGTTTGA 11230

Qy    2204 CACGAGCATGTATCATTCAGTGAGCATATACCTCCAGCGAGGAAAGCAAGAGTTTGAG 2263
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11229 CACGAGCATGTATCATTTAGTGAGCATATACCTCCAGCGAGGAAAGCAAGAGTTTGAG 11170

Qy    2264 CGAAGCTCACAGGAACAACAGCAAGCTGCGAGAAGTCGTGCAAGAAACAGGATAAGTTACT 2323
           |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db   11169 CGAATCTCACAGGAACAATAGCAAGCTGCGAGAAGTGGTGCAAGAAACAGGATAAGTTACT 11110

Qy    2324 CGCCAAGTGCCTCAGGGCTATCAGTTTCTGAAGGACAAGATCAGCTGCTCCTCTTCCAC 2383
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11109 CGCCAAATGCCTCAGGGCTATCCAGTTTCTGAAGGACAAGATCAGCTGCTCCTCTTCCAC 11050

Qy    2384 TACAACTATTCCGCAATGACAGCTCCCTCAGGACATGCCTTCGAGGAGATGACGGCGCC 2443
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11049 TACAGCTATTCCGCAGGGACACACTCCCTCCCGACATGCCCTCGAGGAGATGACGGCGCC 10990

Qy    2444 CGAGCC------------TAGAGACAGAAGATTCTGCATGTCCCTGCGAG 2482
           |||||              ||||||||||||||||||||||||||||||
Db   10989 TGAGCCAAGTCGGCCGCAGGCCTGAGCATAGCAGGCCGAGATTTCGCATGTCCCTGCGAG 10930

Qy    2483 GCATTCATCATTCGAGCCTCGTGAATCTAGGAAGAATCAGGAGAATGCACACTCACTCGATC 2542
           ||||||||||||||||||||||||   ||| |||||||||||||||||||||||||||
Db   10929 GCATTCTTTATTCGAGCCTCGTGAATCAGGAGCTGAGAGTGAAGGAGGACCCACACTCACTCGATC 10870

Qy    2543 TAGCAGCAGGAGCAGACGACTTCTTGCAGTCTCGTAGTTTACGGCGACCGGGTGCTGGCCG 2602
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10869 TAGCAGCAAGAGCAGACGACGACTTCTGCAGTCTTGTAGTTTACGGCGACCGGGTGCTGGCCG 10810

Qy    2603 CAATAGAAGAGAGAGCGAGTATCCTCGAGTATCCTCAGAGCGGTGCTGGCCGGTCGAGAGTGATGA 2662
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10809 CAGAAGAAGAGAGAGGTGCAGGTCCGAGTATCCTCAGAGCCGTGTTGGCCGCCACAGAGCTGATGA 10750

Qy    2663 GATCCGAGTACCACCATGCTGGAGCTGATACGGAACATGGCGGTTCGTCTATGGCTTGGGA 2722
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10749 GGTCGAGTACCCACCCTGCTGGAGCTGATACAGAACAGAGGGCAGTTCGTCTATGGCCTGGGA 10690

Qy    2723 GCAATCACAGGCAGCCATTGACTACCAACTTCGTTCATTATTCGACTGAGGTAAGCGCCT 2782
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10689 GCAATCACAGGCAGCCATTGACGACCAACTTCGTTCATTCTTCGACTGAGGTAAGCGCCT 10630

Qy    2783 CACTTCACCATTCACCATTGTTATTATATCATCTCTTGGATTTGTTCTTTATTTGTTTGTTATTTCAGTGATT 2842
           ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   10629 CACTTCACCATTCACCATTGATTATCATCTCTTGTAATTTGTTCTTTATTTGTTTGTTGTTTTGATT 10570
```

```
Db    2843 GGATTTGTCCTGAGTACTCTTCCAAGTTTATTCACACAGTGACTGTGTGATTAAGT 2902
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10569 GGATTTGTCCGAGTACTCTCTTCCAAGTTTATTCACACAGTGACTGTGTGATTTAAGT 10510

Qy    2903 TTGGGGAGGGCTCAGGAA--GTATGTTGCATTGTATATATTTTTAAGTCTGCATTCATC 2960
           |||||||||||||||||  |||||||||||||||| |||||||||||||||||||||
Db   10509 TTGGGGAGGGCTCAGGAGTGTGTGTTGCATTGTATATACTCTTGAGTCTGCATTCATC 10450

Qy    2961 TAAGGCATAGAAAAACCAAAAAACAAAAATTAAAAATTCAGAAAATGAATTCACAAAAAAG 3020
           |||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   10449 TAAGGCATAGAAAAACCAAAAAAATTGAAAAATTAAGAAAAATGATTCACARAAAAAA 10390

Qy    3021 AGTGTTCATGTAGTTGCATTACATTAGGATCAAGTCTAGAGTGTTTCATTTAGGATTGT 3080
           |||||||||||||||||||||| |||||||||||||||||||| |||||||||||
Db   10389 AATGTTTATGTAGTTGCATTACATTAGGATGCAGTCTAGAGTGTTCATTTTTGGATTGT 10330

Qy    3081 TGCATATGCATAGGGGATAATGATGAGATAGCCTTGTAAGCA 3122
           ||||||||||||||||||||||||||||||||||||||||||
Db   10329 TGCATATGCATAGGGGATAATGATGAGATAGCCTTGTAAGCA 10288
```

RESULT 12
AC073433
LOCUS   AC073433   117055 bp   DNA   linear   PLN 19-JAN-2001
DEFINITION   Arabidopsis thaliana chromosome 1 BAC F19C17 genomic sequence,
               complete sequence.
ACCESSION   AC073433
VERSION   AC073433.9 GI:12321079
KEYWORDS   HTG.
SOURCE   Arabidopsis thaliana (thale cress)
  ORGANISM   Arabidopsis thaliana
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
               rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 117055)
  AUTHORS   Lin,X., Kaul,S., Town,C.D., Benito,M.I., Creasy,T.H., Haas,B.J.,
               Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
               Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
  TITLE   Arabidopsis thaliana chromosome 1 BAC F19C17 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 117055)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE   Direct Submission
  JOURNAL   Submitted (17-JUN-2000) The Institute for Genomic Research, 9712
               Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
REFERENCE   3 (bases 1 to 117055)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE   Direct Submission
  JOURNAL   Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
               Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
               On Jan 19, 2001 this sequence version replaced gi:12280781.
               Address all correspondence to:at@tigr.org
COMMENT   BAC clone F19C17 is from Arabidopsis thaliana chromosome 1
               The orientation of the sequence is from SP6 to T7 end of the BAC
               clone.
               Genes were identified by a combination of several methods: Gene
               prediction programs including Genscan+ (Chris Burge,
               http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky,
               http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant
               of GlimmerM, see Mihaela Pertea,
               http://www.tigr.org/softlab/glimmerm_htm/glimmerm.html, and
               GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
               mpertea@tigr.org), searches of the complete sequence against a
               peptide database and the plant EST database at TIGR
               (http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
               indicate the level of evidence for their annotation. Genes with
               similarity to other proteins are named after the database hits.
               Genes without significant peptide similarity but with EST
               similarity are named as unknown proteins. Genes without protein
               or EST similarity, that are predicted by more than two gene
               prediction programs over most of their length are annotated as

```
                    hypothetical    proteins.  Genes encoding tRNAs are predicted by
                    tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
                    Simple repeats are identified by repeatmasker (Arian Smit,
                    http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES            Location/Qualifiers
     source         1..117055
                    /organism="Arabidopsis thaliana"
                    /mol_type="genomic DNA"
                    /cultivar="Columbia"
                    /db_xref="taxon:3702"
                    /chromosome="I"
                    /clone="F19C17"
     repeat_region  103..397
                    /rpt_family="Rf:centromeric repeat from T7M24 69183 to
                    70294"
     repeat_region  305..399
                    /rpt_family="Rf:ATHILA|ATHILA A.thaliana DNA for Athila
                    retroelement.#Rf:ATR0033|X81801 Athila retroelement"
     repeat_region  305..399
                    /rpt_family="Rf:gi|895704|emb|X89198.1|ATSATDNA4
                    A.thaliana DNA of a 180bp satellite
                    junction#Rf:ATR0037|X89198 clone H33 of Athila 180-bp
                    repeat"
     repeat_region  305..397
                    /rpt_family="Rf:ATHILA|ATHILA A.thaliana DNA for Athila
                    retroelement.#Rf:ATR0033|X81801 Athila retroelement"
     repeat_region  305..397
                    /rpt_family="Rf:Athila LTR for Athila retroelement"
     repeat_region  complement(435..548)
                    /rpt_family="Rf:centromeric repeat from T7M24 67544 to
                    67935"
     repeat_region  complement(1293..1409)
                    /rpt_family="Rf:278A centromeric repeat X92081, found in
                    centromeric regions, 20
                    copies#Rf:gi|1030727|emb|X92081.1|ATRBP278A A.thaliana
                    repeat region (clone 278A) #Rf:ATR0023|X92081 278A"
     gene           2289..3745
                    /gene="F19C17.24"
                    /note="pseudogene, putative retroelement pol polyprotein"
                    /pseudo
     repeat_region  3899..4080
                    /rpt_family="Rf:gi|7330787|gb|AF204875.1|AF204875
                    Arabidopsis thaliana conserved centromere sequence CCS2"
     repeat_region  4135..4873
                    /rpt_family="Rf:gi|7330786|gb|AF204874.1|AF204874
                    Arabidopsis thaliana conserved centromere sequence CCS1"
     repeat_region  4212..4354
                    /rpt_family="Rf:gi|7330787|gb|AF204875.1|AF204875
                    Arabidopsis thaliana conserved centromere sequence CCS2"
     gene           4878..5527
                    /gene="F19C17.25"
                    /note="pseudogene, putative retroelement pol polyprotein"
                    /pseudo
     repeat_region  complement(6987..7083)
                    /rpt_family="(GAA)n"
     repeat_region  complement(7028..7084)
                    /rpt_family="(GAAAA)n"
     gene           7414..8053
                    /gene="F19C17.26"
                    /note="pseudogene, putative retroelement pol polyprotein"
                    /pseudo
     repeat_region  7597..7788
                    /rpt_family="Rf:ATR0038|X13291 Ta1 transposon"
     repeat_region  7597..7788
                    /rpt_family="Rf:TA1_AT|TA1_AT A.thaliana (Landsberg) LTR
                    retrotransposon Ta1-1 DNA."
     repeat_region  complement(8636..9063)
                    /rpt_family="Rf:ARNOLD3|ARNOLD3 Autonomous DNA-transposon
                    ARNOLD3."
     repeat_region  complement(8636..9063)
                    /rpt_family="Rf:ARNOLD4|ARNOLD4 Autonomous DNA-transposon
                    ARNOLD4."
     repeat_region  complement(8636..8980)
```

```
                    /rpt_family="Rf:ARNOLD2|ARNOLD2 Autonomous DNA-transposon
                    ARNOLD2."
     repeat_region  complement(8636..8980)
                    /rpt_family="Rf:ARNOLD3|ARNOLD3 Autonomous DNA-transposon
                    ARNOLD3."
     repeat_region  complement(8636..8980)
                    /rpt_family="Rf:ARNOLD4|ARNOLD4 Autonomous DNA-transposon
                    ARNOLD4."
     repeat_region  complement(8636..8878)
                    /rpt_family="Rf:ARNOLD2|ARNOLD2 Autonomous DNA-transposon
                    ARNOLD2."
     repeat_region  complement(8636..8878)
                    /rpt_family="Rf:ARNOLD4|ARNOLD4 Autonomous DNA-transposon
                    ARNOLD4."
     repeat_region  complement(8636..8875)
                    /rpt_family="Rf:ARNOLD1|ARNOLD1 Autonomous DNA-transposon
                    ARNOLD1."
     repeat_region  complement(8636..8875)
                    /rpt_family="Rf:ARNOLD3|ARNOLD3 Autonomous DNA-transposon
                    ARNOLD3."
     repeat_region  complement(8640..8888)
                    /rpt_family="Rf:ARNOLD2|ARNOLD2 Autonomous DNA-transposon
                    ARNOLD2."
     repeat_region  complement(8640..8883)
                    /rpt_family="Rf:ARNOLD1|ARNOLD1 Autonomous DNA-transposon
                    ARNOLD1."
     repeat_region  complement(8640..8812)
                    /rpt_family="Rf:ARNOLD3|ARNOLD3 Autonomous DNA-transposon
                    ARNOLD3."
     repeat_region  complement(8641..9231)
                    /rpt_family="Rf:ARNOLDY1|ARNOLDY1 Non-autonomous DNA
                    transposon ARNOLDY1 - a consensus sequence."
     repeat_region  complement(8641..8809)
                    /rpt_family="Rf:ARNOLD4|ARNOLD4 Autonomous DNA-transposon
                    ARNOLD4."
     repeat_region  complement(8660..8807)
                    /rpt_family="Rf:ARNOLDY1|ARNOLDY1 Non-autonomous DNA
                    transposon ARNOLDY1 - a consensus sequence."
     repeat_region  complement(8667..24255)
                    /rpt_family="Rf:ARNOLD3|ARNOLD3 Autonomous DNA-transposon
                    ARNOLD3."
     repeat_region  complement(8672..9425)
                    /rpt_family="Rf:ARNOLD4|ARNOLD4 Autonomous DNA-transposon
                    ARNOLD4."
     repeat_region  complement(8681..8795)
                    /rpt_family="Rf:ARNOLD2|ARNOLD2 Autonomous DNA-transposon
                    ARNOLD2."
     repeat_region  complement(8683..8795)
                    /rpt_family="Rf:ARNOLD1|ARNOLD1 Autonomous DNA-transposon
                    ARNOLD1."
     repeat_region  complement(8684..8805)
                    /rpt_family="Rf:ARNOLD4|ARNOLD4 Autonomous DNA-transposon
                    ARNOLD4."
     repeat_region  complement(8685..9190)
                    /rpt_family="Rf:ARNOLDY1|ARNOLDY1 Non-autonomous DNA
                    transposon ARNOLDY1 - a consensus sequence."
     repeat_region  complement(8685..9148)
                    /rpt_family="Rf:ARNOLD4|ARNOLD4 Autonomous DNA-transposon
                    ARNOLD4."
     repeat_region  complement(8685..8883)
                    /rpt_family="Rf:ARNOLD3|ARNOLD3 Autonomous DNA-transposon
                    ARNOLD3."
     repeat_region  complement(8686..8771)
                    /rpt_family="Rf:ARNOLDY1|ARNOLDY1 Non-autonomous DNA
                    transposon ARNOLDY1 - a consensus sequence."
     repeat_region  complement(8686..8765)
                    /rpt_family="Rf:ARNOLDY1|ARNOLDY1 Non-autonomous DNA
                    transposon ARNOLDY1 - a consensus sequence."
     repeat_region  complement(8687..8875)
                    /rpt_family="Rf:ARNOLD1|ARNOLD1 Autonomous DNA-transposon
                    ARNOLD1."
     repeat_region  complement(8688..8809)
                    /rpt_family="Rf:ARNOLD2|ARNOLD2 Autonomous DNA-transposon
```

```
repeat_region    ARNOLD2."
                 complement(8689..8849)
                 /rpt_family="Rf.ARNOLD1|ARNOLD1 Autonomous DNA-transposon
                 ARNOLD1."
repeat_region    complement(8689..8849)

Query Match      61.2%;  Score 1911.6;  DB 8;  Length 117055;
Best Local Similarity  90.1%;  Pred. No. 0;
Matches 2147;  Conservative  0;  Mismatches  174;  Indels  61;  Gaps  7;
```

```
Qy    796  GGTTTACTCGACCGGTACGCTGGTTAGATTGTGTTTTGGTTTTGCTATTACTAACATA   855
           ||||||||||||||||||| ||||||||||||||||| || |||||||||||||||||
Db 112558  GGTATACTCGACCGATTACTCTGTTGGTTTGTGTTTTGGTTTTGGTTAGTTGGGACTAACCGTA 112617

Qy    856  TTAACGTTTATCTTGAGTTCGCTCGTTTTTAGGTTTCATCATGAGTAACTACAGTGGA   915
           |||||||||||||||| |||||||||||| |||||||||||||||||||||||||||||
Db 112618  TTAACGTTTATCTTGAGTTTATCTTTCAGGTTTTCATTGTTTTCATTATGAGTAACTACAATGGA 112677

Qy    916  AAATCCTCTATGGACCCTGATTATATAATGTGAAGCTAAGTCCTGGTCCACTAGACCG   975
           |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db 112678  GAATCCTCCAATGGATCCTGATTATATAATGTAAACGAAGCTGAGTCTTGGT-CACTAGACCG 112736

Qy    976  GAGTGAGAGCACACATGTTACGAGAGCTATAGGAGGGATGAATTTGAACCGCTTCGCAGTCGA  1035
           |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db 112737  GAGAGAGAGCAACATGTTTACGAGAGCTTCAAGGATGAGTACTAGGATTCTCGAGGGCTACTAGGATGAATTTGAACGCTCCGCAGTCGA 112796

Qy   1036  CGTAATCAAAGAAGAGCTGAAATCGCTAGAGGAAAGAGGGCCGATGTCGAGTAGATATGAG   1095
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 112797  CGTAATCAGAGAAGAGCTGAAATCGCTATAGCGATATCGAGTATCGAGTAGTCGAGTATATGAG 112856

Qy   1096  CTGATTGATGAGGATATCAAAACTGAGTATGAGCCAGAGTCATGGCGCAAGGAGAACGAAAG   1155
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 112857  CTGATTGATGAGGATATCGAAACTGAGTATCGAGTCCAGAGTCATGGCACAAGGAGAACGAAAG 112916

Qy   1156  CTACTGAACAAAATCCGACGAGAGGTTACAGTGGAGTGAGGAGTATATCAGATTCTTTGAGATGAAT   1215
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 112917  CTGCTGAACAAACCCGACGAGAGGTTACAGTGGAGTGAGGAGTATGTCAGATTCTTTGAGATGAAC 112976

Qy   1216  GACTTCTGGGAACGAGCAGCGTATCCCTGATAATGAGACTTTAGCCCAGTTGGGGTTACTGGAG   1275
           ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db 112977  GACTTCTGGGGAACGAGCAGCGTATCCCTGATAATGAGACTTTAGCAGCTGGAGTGGGGTTACTGGAG 113036

Qy   1276  GACGTGCAGCATCTGTTCGAGAAGTGCATCTG-------------------------------------------------   1308
           ||||||||||||||||||||||||||| ||
Db 113037  GACGTGCAGCATCAATTGAGAAGTGTCATCTGGAGACACTCATGTCTTACCCTTATGCT 113096

Qy   1309  ---ATAAGGAGGAACAATCGAGTTTCTTCCACCACTGGGAGTTCCACCACTGCGAAGTGTATGAGGGA   1364
              |||||||||||||||||||||| |||||||||||||||||||||||||
Db 113097  GCCTATAAGGAGAACCATCGAGTTTGGGGTCGTTAGGGTTAGGGTTAGTTCAGTGTATGAGGGT 113152

Qy   1365  CTCACAGAGCTTGAGCTGGATACCATGGGGTTAGGCTTCTTGACGTTCTTAGTGATGAA   1424
           |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db 113153  CTCACGGAGCTTGAGCTGGATATCGAGTTTCTTGACGTTCTTCAGTGATGAA 113212

Qy   1425  CAGCGGGTACCAGTTTAGATCAAGAAATTGGAAGAACTGTTTGGTTTCCCTAGTTGGAAAG   1484
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113213  CAGCGGGTATCCAGTTTGGATCAAGAAGTGTTTGGTTTCACTAGTGGAAAG 113272

Qy   1485  GGAACCAACCCAGTTGACAGGGAAGAGCTTAAGGACTTAAGGACGTAGGGTGGGCTACTATTGGGAAC   1544
           |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db 113273  GGAACCAAACCCAGTTTGACAAGGAAGAGCTTAAGGACGTAAGGTGTGGGCTACTATTGGGAAC 113332

Qy   1545  AAITCTACCGCTAAACTGACGCGGTCCAAGAGCAACCAAATCCGAGTCCGTGATTCGC   1604
              |||| |||||||||| ||||||||||||||| ||||||||||||||
Db 113333  AAT-TATCGCTAAACTCGGCACGGTCCAAAGCAACCAAATCCGAGTCCGTGATCCGC 113391

Qy   1605  TACTTTCAGCGCTCGGTTGCCAATGTTTTTACTTCCACAGGCACCGTGTCT   1664
           ||||||||||||||||||||||||||||| ||||||||||||||||||||
Db 113392  TACTTTCAGTGCTCGGTTGCCAAGCTTTTCTACTTCCACAGGCACCGTGTCT 113451

Qy   1665  AACACAGACGAAGAGATGAGATGATAGATTCAGCGTTTATAGGGATTCTCCGCCTTACAAAAGGA   1724
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113452  AACACGGACAATGGAGATGATAGAGATAGATTTCGCGGATTCTTCCGCCTTACAAAAGGA 113511
```

```
Qy   1725  AAGAAATGTCCTGAGAGGAGATCTTAACGACTCACCACCAGTAATGCTTCTTGTTGATCCAT   1784
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113512  AAGAAATGTCTTGAGAGGGCGATCTCAACACGCACCAGTAATGCTTCTTGTTGATCCAC 113571

Qy   1785  CTGTGTGGGTACATGAAGTGGGCGTGACAAACGGCAAGAAGAAGGTAAGAGG-AGCACT   1843
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113572  CTGTGTGGATACAGGAGGCGGCGTTGATAAAGGCCAAGAAGGTGAGGAGCCAGCACT 113631

Qy   1844  ATGCGTGGGTGGCGTTGTGACGCCAATTCGAAAGTTTGTGGAGTTCCGCTCAAGGAAGT   1903
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113632  ATGCGTGGGTGGCGTTGTGATGCGAATTGGGAGCCAATTTTGGAGCTTGTGTGGAGTTCCGCTCAAGGAGCC 113691

Qy   1904  AGGGTTAGCACCGAGAATGATGAGCTTGGCCGATCGTGAGTTCTTGAGTT   1963
           |||||||||||||||||| ||||||||||||||||||||||||||||
Db 113692  AGGGTTAGCACCGAGAATGATGATCTTGTGCCGTGATGTGAGTTCTTGGAGTT 113751

Qy   1964  TGACATGGTTGGGCGACTTTCACCGGTTCGAGCATTCATCGATTAGAATCGCCAA   2023
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113752  TGACATGGTTGGGCGACTTTCACCGGTTCGAGCATTCATCGATTCGATTCGCCAA 113811

Qy   2024  CATTCTTTTCCCCTGCATTTACGTACTAGGATTCTCGAGGGCAGGAACATTGACTTCAA   2083
           ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db 113812  CATTCTTCTCCCCCTGCATTTACGTACTAGGATTCTCGAGCAAAAGGAACATTGACTTCAA 113871

Qy   2084  GCCTGCGCTTGAAGATCTTTATTTGAGGGGCAGTCCGCCAACTGAGAGATTAGTCACAC   2143
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113872  GCCTGCGCTTGAAGATCTTTATTTGAGGGCAGTCCGCGACTGAGAGATTAGTCACAC 113931

Qy   2144  CGAAGGAGGAGCTACAATGGAGATTGTTGATGAAGACATATGATAGTAGAGGAGGAGGTTTGA   2203
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113932  CGAAGGAGGACCTACAAAAGAAGATTGTTGATGAACAGATGATGAATGAGATGAGGCGTGTTTGA 113991

Qy   2204  CACGAGCATGTATCATTTCAGTGAGCATGATAATCAGATTCTTTGAGATGAAT   2263
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113992  CACGAGCATGTATCATTTTAGTGAGCATGATAGTTCAGATTCTTTGAGATGAG 114051

Qy   2264  CGAAGCTCACAGGAACAACAGCAAGTGTGCAAGAGTGAAGATAAGAAGATAAGTTACT   2323
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114052  CGAATCTCACAGGAACAATAGCAAGTGTGCAAGAGTGTGCAAGAAACAGGATAAGTTACT 114111

Qy   2324  CGCCAAGTGCCTCAGGCTATCGAGGACAAGATCAGCTGCTCCTTCCAC   2383
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114112  CGCCAAGTGCCCAAATGGCTATCGAGGACAAGATCAGCTGCTCCTTCCAC 114171

Qy   2384  TACAACTATTCGCAAGTTCTTCCACCACTGCGAAGTGTATGACGCGCC   2443
           ||||||||||||||| |||||||||||||||||||||||
Db 114172  TACAGCTATTCGCAAGTTCTTCCACCACTCCCGGGTTCCGAGATATGACCGCGCC 114231

Qy   2444  CGAGCC---------------------------------TAGAGAGCAGAAGATTCTGCATGTCCCTGCGAG   2482
           |||||
Db 114232  TGAGCCAAGTCGGGCAGGGCCTCGAGCAGGGGGCCCAGGCATAGAGAGCAGAGATTCGCATGTCCCTGCGAG 114291

Qy   2483  GCATTCATCATTCCGAGCCTCGTGAATCGAGCCTCAGTGGATGAA   2542
           ||||||||||||||||||||||||||||||||||||||||||||
Db 114292  GCATTCTTTATTCGAGCCTCGTGAATCGAGCCTCACTAGTGGATGAA 114351

Qy   2543  TAGCAGCAGGAGCAGACGACTTCTGCGAGTCTCGTAGTTTACGCGACGTGCTCGCCG   2602
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114352  TAGCAGCAAGACCAGCGACGACTTCTGCGAGTCTTGTAGTTTACGGACGTGCTCGCCG 114411

Qy   2603  CAATAGAAGAAGAGGTCGAGTATCCTCAGAGCGCCGTGCCCGCGCCACAGAGCTGATGA   2662
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114412  CAGAGAAGAAGAGGTCGAGTATCCTCAGAGCGCGTGTTGGCCGCGCCACAGAGCTGATGA 114471

Qy   2663  GATCGAGTACCCACATGCTCGGAGCTTGGAGCGGTTCGTCTATGGCTTGGGA   2722
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114472  GGTCGAGTACCCACCTGCTCGGAGCTTGGAGCGGTTCGTCATGGCCTATGGCCTGGGA 114531

Qy   2723  GCAATCACAGGCAGCCAGCCTCTCCGGCCTTACAAAAGGA   2782
           ||||||||||||||||||||||||||||||||||||||||
Db 114532  GCAATCACAGGACCAGCCAGCCAACTTCGTTCGTTCGACTGAGCTGAGGTAAGCGCCT 114591
```

```
Qy   2783  CACTTCACCATTATATATATCATCTCCTGTGATTGTTCTTTTATTTTGTTTCAGTGATT 2842
                  ||||||||||||||||||||| ||||||  || |||||||| ||||| |||||
Db 114592  CACTTCACCATTGTATTATCATCATCTCTGTAATTTGTTCTTTTATTTTGTTTTGTGATT 114651

Qy   2843  GGATTGTCCGAGTACTCTCTTCCAAGTTATTCACCACAGTGGACTGTGTGATTTAAGT 2902
                  |||||| |||||||||||||||||||||||| |||||||||| |||||||||||||
Db 114652  GGATTGTCCCGAGTACTCTCTTCCAAGTTATTCACATAGTGGACTGTGTGATTTAAGT 114711

Qy   2903  TTGGGGAGGGGCTCAGGAA--GTATGTGCATTGTATATATTTTAAGTCTGCATTCATC 2960
                  ||||||||||||||||||  ||| ||||||||||||||||| ||||||||||||||
Db 114712  TTGGGGAGGGGCTCAGGAAGTGTCGTTGCATTGTATATATTTTGAGTCTGCATTCATC 114771

Qy   2961  TAAGGCATAGAAAAACCAAAAAATATAAAAATTCAGAAAATGATTCACAAAAAAAG 3020
                  ||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db 114772  TAAGGCATAGAAAAACCAAAAAATTTTGAAAAATTAAGAAAAATGATTCACAAAAAAAA 114831

Qy   3021  AGTGTTCATGTAGTTGCATTACATTTAGGATCAAGTCTAGAGTGTTTCATTTAGGATTGT 3080
                  ||||| |||||||||||||||||||||||||||||||||||||| |||||||||||||
Db 114832  AATGTTTATGTAGTTGCATTACATTTAGGATCGAGATCTCAGAGTGTTTCATTTTGGATTGT 114891

Qy   3081  TGCATATGCATAGGGGATAATGATGAGATAGCCTTGTAAGCA 3122
                  |||||||||||||||||||||||||||||||||||||||||
Db 114892  TGCATATGCATAGGGGATAAATGATGAGATAGCCTTGTAAGCA 114933
```

```
RESULT 13
TSH22/c
LOCUS       TSH22              69752 bp    DNA     linear   PLN 06-OCT-1998
DEFINITION  Arabidopsis thaliana BAC T5H22.
ACCESSION   AF096372
VERSION     AF096372.1  GI:3695394
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 69752)
  AUTHORS   Washington University Genome Sequencing Center.
  TITLE     The A. thaliana Genome Sequencing Project
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 69752)
  AUTHORS   Graves,T., Harmon,G. and Duckels,G.
  TITLE     The sequence of A. thaliana T5H22
  JOURNAL   Unpublished (1998)
REFERENCE   3  (bases 1 to 69752)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     Submitted by:
            Genome Sequencing Center
            Department of Genetics, Washington University,
            St. Louis, MO 63108, USA
            e-mail: rwilson@watson.wustl.edu

            MAPPING: Clones were assigned to the YAC map by hybridization by
            M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted
            by M. Marra, WashU, to pick the best candidates for sequencing.

            NOTICE: This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone
```

NEIGHBORING COSMID INFORMATION:

The 3' clone is T5L23, 200 bp overlap. Actual start of this clone
is at base position 1 of T5H22; actual end is at 69558 of T5H22.

The submitted sequence of T5H22 has a tandem repeat from base
position 21900 to 6374 . The assembly of the sequence outside of
the tandem agrees with restriction digest information. The
sequence within the tandem is a representation of the region and
may contain misassemblies and single stranded regions.

NOTES:

Coding sequences below are predicted from computer analysis, using
the program Genefinder(P. Green and L. Hillier, ms in preparation).

```
FEATURES             Location/Qualifiers
     source          1..69752
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="IV"
                     /map="unknown"
                     /clone="T5H22"
     gene            complement(550..2361)
                     /gene="T5H22.2"
     CDS             complement(join(550..856,1274..2361))
                     /gene="T5H22.2"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAC62798.1"
                     /db_xref="GI:3695397"
                     /translation="MVNNQIPGNTVDVEGNPIPPIQTDVPEAATPVTLAELRSMMAQL
                     QQKVNDQEQANRSLAQQLEAATSQGQIRTTRFGARHLQDRRAAADLNPTRFVFHTPGN
                     TTRPVRRTAPEIGRDRTEPAILGNRETKRTERNKPQLPPRAEVTEADHIGVSDDKDL
                     EENIRWAEEYATEQEISAIKLSLAKAENEMKLVRSQMHNAVSSAPNIDRILEESHNTP
                     FTHMISNAIISDPGKLRIEYFNGSSDPKGHLKLFIISVARAKFRPEERDAGLCHLFVE
                     HLKGPALNWFTRLKGNSVDSFQELSTLFLKQYSVLIDPSTSDADLWSLSQQPNEPLRD
                     FLAKFRSTLAKVEGINDLAALSTLKKALCRARLSAKFLAREIGGGLTIKDLEAEKGK
                     TEQVNAVANPEQAAPAANSEGPKRGRGNREADDDEPEAARGRIFTVLGDSAFCQDTAA
                     SIKAYQRKADANHN"
     gene            5708..5965
                     /gene="T5H22.1"
     CDS             join(5708..5839,5894..5965)
                     /gene="T5H22.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAC62799.1"
                     /db_xref="GI:3695398"
                     /translation="MVQKVLELLHAKKVLGTSSMGRELCSGSLTGACACLSLRKRDSK
                     KEEVQRKEIERDMSIRDCPGMGG"
     gene            complement(6472..8565)
                     /gene="T5H22.3"
     CDS             complement(join(6472..7313,7380..7984,8543..8565))
                     /gene="T5H22.3"
                     /note="contains similarity to Arabidopsis thaliana
                     retrotransposon Athila (GB:X81801)"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAC62797.1"
                     /db_xref="GI:3695396"
                     /translation="MVIESGTSYQQPNPPGFTQQPQAPSAQDPEIKQLIQQIIRGQ
                     ATGAMAFDKIAELNKIDCSVNDLNVKFEALNSKIKYVESQFASTSAPKHPQQLPRK
                     AVQNPKDYANANAITIHQEDESPPSHRTPNADNMIQEGGDTQIAAPATQQFIWTIE
                     HTPLLNTHHFGKFKFIKKLKKDQNQEMISRVSTQETYKKKVIQEKLDDPASVSLMPLS
                     TAKRLGIMEYKPCNLALLLADVSVAHPHGLAQTEDPTEDDWYEMRKKFKWQSRVIEGF
                     SSRVMKLKDQLWMFEKRVNKFSTGIDPEKMETTLPIKEDEDFTREWFAPEDQKGVHLE
                     ERSLESELSEEBISIPIYSPGBPKPPTNPSAEEEQLILDSFAEKLLEIWIGYNAQED
                     TIGKSTHKMEELNSTILELQEMVKGYHNFEIEDWSKEPGSEEEDYSTNQKEAYFEEKS
                     NEYSSYHMSREDAEYDSCTENSTTEDEDFYVPLFNPFSD"
     gene
```

```
CDS          /gene="T5H22.4"
             complement(join(11247. .12050,12376. .12638,13526. .13854,
             14387. .14590,14707. .15866,16465. .16899))
             /gene="T5H22.4"
             /note="contains similarity to Arabidopsis thaliana
             retrotransposon Athila (GB:X81801)"
             /codon_start=1
             /evidence=not_experimental
             /protein_id="AAC62800.1"
             /db_xref="GI:3695399"
             /translation="MADEQNQONGPTNVGACDAPRDHRQRKGIAPPAIQNNNFEIKSG
             LISIIQGNKFHGLPMEDPLDHLDEFDRLCNLTKINGVSEDGFKLRLFPFSLGDKAHIW
             EKNLSHDSITTWMIARRLFYQSSSMPELQDSEMRCLAFVRRLGNYQPQPHGFAPQE
             NQGFTAPDAEMKQMVQQLLQQQASSSMEIPKKLSELHHKLDCNYNDLNAKVEALNTKV
             RYLEGQSASTSVPKVTGLPGKSIQNPKEYATAHAITICHDRELPTRPVLDLIIGESDV
             QEGEACTQIEVSVIEFNHSAGSHHLIQSTSBEKAAIIERMVKRFKPTLPSRALPWTF
             RKAWIERYKSVAAKQLDEIEAVMPLMEVLNLIPDHHKDVRNLILEKIKMVHDSDDESD
             ATPSRVDKRIVQEKLEDPGSFTLPCSIGELAFSDFLCDLGASVSLMPLSMGRRLEFI
             HYKPCDLTLLADRSSRKPGMLKDLPVMLNGVEVPTDFVVLNMEVKHKDPLILGRPF
             LASVGAVIDIREGKISLNLGKHIKSDTQEETIERLAHSVELRSKINQMOKEAQPKGG
             INTIPRKKFTSRWSEETDYPLEEKEAYFEERGIENPTPSSHNHSTPSYPISTHHSTS
             TPSPSLTYHLLDLVISSTPIKLYPIPPHSSSPTNQLSYSQOPLDLLTIFFIYSTSTSFS
             THQTSSSVRFHHSIAAAFVSLPPPPLIHLWATIGNNLPLNLARSKSNQIQSPVICYFQ
             RSVANVFYSRESTSVSNTDMEMIDSALTGILRRTKGKNVLRGDLNNAPEMSVDPPG
             SPPTDDVSHTEGAPTEDVDETDDIDEAEFNTSMVHFSEHIPPARESKSLSEAHRKNSK
             LQKWCKKQDKLLVKCLRAIKFLKDKISCSSSTTAIPQGQLSQDMPSRRYDAPEPSRRR
             PEPRROEIPHVPARHSSFEPRESGRRGPHSLDLAEGADDFYSLVYYATAVLTAVEEE
             RSSILRAVLAATELMRSSTHLLELIRNKAVRLWPGSNHMQPLTTNSVHSSTKVSASLH
             HCIISSVVICSLFCFCDWICPVYSLPSLFTQWSV"

gene         complement(66837. .69653)
             /gene="T5H22.5"
CDS          complement(join(66837. .67381,67427. .67843,68078. .68338,
             6838. .68841,69424. .69653))
             /gene="T5H22.5"
             /note="contains similarity to reverse transcriptase (Pfam:
             PF00078 rvt, E=4.3e-08)"
             /codon_start=1
             /evidence=not_experimental
             /protein_id="AAC62796.1"
             /db_xref="GI:3695395"
             /translation="MRSKFHQATSSEPDINRVIEARRTIFIPQIASLRIRNSRKLNL
             EPYNGLEDPKGYLAAFLIAAGWVDLDEAEEDARVTEHLTKDCIVLKRYLAELWASGDL
             SKFNIDDFVKQYHEARDNSETQNSKRPRLTNEEBPRSSKGKINVILGGSKLCHNSVSA
             IKKHRPNVLLKSSLSEEVNYQCSSVSFDEEETRHIERPHDDALIITLDVANFKISRIL
             VDTGSSVDLIFLGPPSPLVAFTSESAMSLGTIKLPVLAGKMSKIVDFVFVFDKPATYNI
             ILGTPWIFQMKAVPSTYHOWLKFPTSNGVETIWGDLEGSRTSAYMPDIDPSIICHELN
             VDPRFKPLKQKRRKLGVERAKAVNGKIDKLLKIGSIREVQYPDWVAITVVVKKNGKD
             RVCIDFTDLNKACPKDSFPLPHIDRLVESTAGNELLTFMDAFLGYNQIMMNPBDQEKT
             SFITDRGATYQWLVNKMFNEHLRKTMEVSIDDTLVKSLKKEDHVKGHLGECFEILNQYQ
             RSTDKSLPFYQILKGNNGFIWDEKCEEAFRQLKAYLTTPPVLSKPEADEKLYLYVFVS
             NHAVSGVLVREDR"

ORIGIN

Query Match      61.2%;  Score 1910.6;  DB 8;  Length 69752;
Best Local Similarity  90.1%;  Pred. No. 0;
Matches 2146;  Conservative  0;  Mismatches  174;  Indels  61;  Gaps  7;
```

```
Qy    1036  CGTAATCAAAGAAGAAGAGCTCTAGAGGAAGAAAGAGAGGCGATGTCGAGTAGATATGAG  1095
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13155  CGTAATCAGAGAAGAGTGCTGAAATCGCCAGAGAAGAAAGAGAGGCGATGTCGAGTAGATATGAG  13096

Qy    1096  CTGATTGATGAGGATATCAAAACTGAGTATGAGCCAGAGTATGCGCAGAGAGACGAAG  1155
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13095  CTGATTGATGAGGACATCGGAACTGAGTATGAGCAAGAGTCATGCGCAGAGAGACAAAG  13036

Qy    1156  CTACTGAACAAATCCGACGAGGTTACAGTGGAGGAGTATATCAGATTCTTGAGATGAAT  1215
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13035  CTGCGGAACAAACCTGACGAGGTTACAGTGGAGGAGTATATCAGATTCTTTAAGATGAAT  12976

Qy    1216  GACTTCTGGGGACCGACGGTATCCCTGATATGGAGACTTTAGCCCAGTTGGGGTTACTGGGAG  1275
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   12975  GACTTCTGGGGAACGACGAGGTATTCGTCGTATGGAGACTTTAGCCCAGCTGGGGTTACTAGAG  12916

Qy    1276  GAGCTGCAGCATCGTCGTTGGAGAAGTGTCATCTG--------------------------  1308
            ||||||||||||||||||||||||||||||||||
Db   12915  GAGCTGCAGCATCGTCGTTGCGATCATCTGGAGACACTCAGTGTCTTACCCTTACGCT  12856

Qy    1309  ---ATAAGGAGGAACAATGGAGTTTCTTTCCCACACTGCAGTGGAAATGTATGAGGGA  1364
               ||||||||||||||||||||||||||||||||||||||||||||||||
Db   12855  GCTTATTAGGAGGAACAATGGAGTTTCTTTCCCACACTTCAAGTGGAGATGTATGAGGGA  12796

Qy    1365  CTCACAGACTTTGAGCTGGATACCATGGGGTTAGGCTTCTTGACGTTCTTAGTTGGATGAA  1424
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   12795  CTCACAGACTTTGAGCTGGATACCATGGGGTTAGGCTTCTTGACGTTCTTAGTTGGATGAA  12736

Qy    1425  CAGCGGTACCAGATTTAGATCAAGATTAGGAAGAAATTGGAAGAACTGTTTCGGTTCCCTAGTGGAAAG  1484
            |||||||||||||||||||||||||||||||||||||||||||||||
Db   12735  CAGCGGTATCAGCTTGCGATCAGAGTTTGACACGGGATCAAGAAGTTGAAGAACTGTTTCGGTTCCCTAGTGGAAAG  12676

Qy    1485  GGAACCAACCCCAGGTTTGACAGAGACTTAAGGATTTGTGGGCTACTATTGGGAAC  1544
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db   12675  GGAACCAAACCCAGATTTGGGAATTTGTGGGCCACTATGGGAGATGGAGATTGAGGGAC  12616

Qy    1545  AATCTACCGCTAAACTGCGACGCGGTCCAAGAGCAACCAAATCCAAGTCCTGATTCGC  1604
            |||||||||||||||||||||||||||||||||||||||||||||||
Db   12615  AATCTACCGCTAAACTTGGCGCGGTCTAAGAGCAACCAAATCCAAGTCCTGTGATCTGC  12556

Qy    1605  TACTTTCAGCGGCTCGGTTGCCAATGTTTTTACTCCAGGGAGTCACAGGCACCGTGTCT  1664
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   12555  TACTTTCAGCGGCTCGGTTGCCAACGTCTTTTACTCCAGGGAGTCACAAGCACCGTGTCT  12496

Qy    1665  AACACAGACATGAAGATGATAGATTCAGCGCTTATAGGGATTCCGGCCTTACAAAAGGA  1724
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   12495  AACACGACATGGAGATGATAGATTCCGGCCTTACAGGGATTCCGGCCTACGAAAGGA  12436

Qy    1725  AAGAATGTCCTGAGAGAGATCTTAACGACTCACCACCACCAGTAATGCCTCGTGTTGATCCAT  1784
            ||||||||||||||||||||||||||||||||||||||      |
Db   12435  AAGAATGTCTTGAGAGAGATCTCAACACACGCACCAGAAATG--TCTGTTGATCCAC  12378

Qy    1785  CTGTGTGGGTACATGAAGTGGGCGTCGACAAGCGCAAGCAAGAGAAGTAAGAGACACTA  1844
            |||||||||||||||||||||||||||||| |
Db   12377  CTGTGTGGGATATAGGAAGTGGGCGTTGGGAGGTTGGGAGTTCCGCTCAAGGAGAGCACTA  12318

Qy    1845  TGGCGTGGGTGGGCGTTGTGACGCAATTCGAAAGTTTGTGGAGTTCCGCTCAAGGAAGTA  1904
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   12317  TGCGTGGGTGGGCGTTGTGATGCCAATTCTGGAAGCTGTGGGAGGATTCCGCTCAAGGAACCA  12258

Qy    1905  GGGTTAGCACCGAGAATGATGGACTTGGATCACTTGCGCCGATGGAACTAACAGTGAA  1964
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db   12257  GGTTTAGCACTGAGGATGGATGACTTGGATATTTGCGCCGATGGAACTAACAGTGAA  12198

Qy    1965  GACATGGTTGGCGACTTTCACCGGTTCGAGCATTCATCGATTAGAATCGCCAGACCG  2024
            ||||||||||||||||||||||||||||||||||||||||||||||
Db   12197  GACATGGCTGGCGACTTTCCACCGCTACAGGTTCGAGCATTCATCGATTAGAATCGCCAAC-  12138

Qy    2025  ATTCTTTTCCCCTGCATTTAGCGTACGAGCTTACCAGGGATGAATGCTTCGACCTCGA  2084
            ||||||||||||||||||||||||||||||||||||||||||||||
Db   12137  ATTCTTCTCCCCTGCATTGAGCACGAGCTTACCAGGGATGAATGCTTCGACCTCGA  12078
```

```
Qy     796  GGTTTACTCGACCGTACGCTCGGTTGAGTTGTGTTTTTGGTTTTGGTCTATTACTAACATA  855
            |||      |                |  ||||||||  |||
Db   13392  GGTACACTCGACCGGTTTACTCTCGTGATTGGTTGTGTTTTGGTTTTGGTTGGAACTAACGTA  13333

Qy     856  TTAACGTTTATTCTTTGAGTTTCGTCTGTTTTTAGGTTTCATCATGATAACTACAGTGAA  915
            |||       ||     |   ||||||||||  ||||
Db   13332  TTAACGTTAATTTTGTTTTGTTTTCAGGTTTCATCATGATAACTACAGTGGA  13273

Qy     916  AAAATCCTCATGGACCCTGATTATATAATGTCGATGAAGCTAAGTCCTGGTCCCACTAGACCG  975
            |||||||||||||||||||||||||||||||||||
Db   13272  GAATCCTTCGGACCCCTGATTATAATGTCGATGAAGCTGCAGTTCAGGTTCCGCTCGACC-  13214

Qy     976  GAGTGAGAGCAACATGTTTACGAGCTACTATAGGGATGAATTGAACGCTCGCAGTCGA  1035
            || ||||||||||||||||||||||||||||||||||||||||||
Db   13213  --GAGAGCAACATGTTTACCAGGAGCTTCAGGGATGAATTTGAATGCTCGCCAACTCGA  13156
```

```
RESULT 14
ATCHRIV9/c
LOCUS           ATCHRIV9              159629 bp    DNA     linear   PLN 16-MAR-2000
DEFINITION      Arabidopsis thaliana DNA chromosome 4, contig fragment No. 9.
ACCESSION       AL161497
VERSION         AL161497.2  GI:7270670
KEYWORDS        .
SOURCE          Arabidopsis thaliana (thale cress)
  ORGANISM      Arabidopsis thaliana
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
                rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE       1  (bases 1 to 53060)
  AUTHORS       Spiegel,L.A., Huang,E.N., Nascimento,L.U., de la Bastide,M.,
                Vil,D.M., Preston,R.R., Matero,A., Shah,R., O'Shaughnessy,A.,
                Rodriguez,M., Shekher,M., Schutz,K., See,L.H., Swaby,I.,
                Habermann,K., Dedhia,N.N., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
  JOURNAL       Unpublished
REFERENCE       2  (bases 21279 to 137726)
  AUTHORS       Zhong,J., Ma,P., Parnell,L.D., Chen,C.N., Chen,E.Y., Mewes,H.W.,
                Lemcke,K. and Mayer,K.F.X.
  JOURNAL       Unpublished
REFERENCE       3  (bases 137729 to 137730)
  AUTHORS       Lamar,B., Stoneking,T., Stumpf,J., Mewes,H.W., Lemcke,K. and
                Mayer,K.F.X.
  JOURNAL       Unpublished
REFERENCE       4  (bases 1 to 159629)
  AUTHORS       EU Arabidopsis sequencing.project.
  TITLE         Direct Submission
  JOURNAL       Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
                Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
                lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project
                Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
                Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
                E-mail: michael.bevan@bbsrc.ac.uk
COMMENT         Information on performance of analysis and a more detailed
                annotation of this entry and other sequences of chromosomes 3, 4
                and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
                this fragment has an overlap with ATCHRIV8 at the 5' end and an
                overlap with ATCHRIV10 at the 3' end.
FEATURES             Location/Qualifiers
     source          1..159629
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /variety="columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     gene            6741..9588
                     /gene="AT4g03480"
                     complement(join(6741..7064,7161..7708,7783..8101,
                     8409..8866,9283..9343,9445..9588))
     gene            /gene="AT4g03480"
                     complement(join(6741..7064,7161..7708,7783..8101,
                     8409..8866,9283..9343,9445..9588))
     CDS             /gene="AT4g03480"
                     /note="contains similarity to mammalian ankyrins
                     similarity to similar to A. thaliana hypothetical protein
                     F21B7.8, GenBank accession number AC002560
                     contains ATP/GTP-binding site motif A (P-loop) AA297-304
                     contains EST gb:H37081, T42092"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB77833.1"
                     /db_xref="GI:7270671"
                     /db_xref="SPTREMBL:Q92T76"
                     /translation="MVRSHLLRATPEASPLKPASSPSNCRLSLSPVELRLYFIRDDDF
                     KVKVDIFKTSYQNSLFELPCCDKSEFLTSLRFSDLFNLPGEYVPMNPEIFSAMRAGNI
                     ELLEKLSYETPMACLKSDGDSVLHLAAASGHLELVNIITECPLLLEPNSKYQIP
                     LHVAARAGRSAVVKALVASVLYFSPRVPEEDRDRLNIYVLKDIDGDTPLHAALKDLHE
                     KABERIRKLSLSHLIMHWRRSRCISFSDASTRQMETAACLVNADQHASFLANKDGTSP
                     LYLAVEAGNSLVRAMLNRPDWIKQIOGKTSTLASQLEGRKSLLHHALAKAXNTDVLNVIL
                     NDDPSLVNERDEEGRTCLSVGASMGYYKGICKLLDRSTKSVYECDKDGSFPIHMAVEK
                     GHIKVVKEILKRCPDSKELVNKQGQNMLHIAAKSAKVGSFLLGYIRRLDTENHLIEEQ
```

SEQUENCE ALIGNMENT (Qy = Query, Db = Database):

```
Qy   2085  CCTGCGCTTGAAGATCTTTATTTCGAGGGCAGTCCGCCAACTGAGGAGTTAGTCACACC  2144
           ||||||||||||||||| ||||||||||||| |||||||||||||||||||||||||||
Db  12077  CCTGCGCTTGAAGATCTTTATTTTGAGGGCAGTCCGCCAACTGAGTGATGATGTTAGTCACACC  12018

Qy   2145  GAAGGAGCTACAATAGAGAGATGTTGATGAGCATATGATATAGATAGAGAGCGGAGTTGAC  2204
           ||||||||| ||| |||||||||||||||||| |||||| ||||||||||||||||||||
Db  12017  GAAGGAGCTCCAACAGAGACAGATGTTGATGAGCACAGATATATAGAGAGCGGAGTTCAAC  11958

Qy   2205  ACAGACATGTATCATTCAGTGAGCATATACCTCCAGCGAGGAAAAGCAAGAGTTTGAGC  2264
           |||| ||||||||||||||| ||||||| ||||||||||||||||||||||||||||||
Db  11957  ACAAGCATGTATCATTCAGTGAGCATATACCTCCAGCGAGGAAAGCAAGAGTTTGAGC  11898

Qy   2265  GAAGCTCACAGGACAACAGCAAGCTGCGCAGAAGTGGTGCAAGAAACAGGATAAGTTACTC  2324
           ||||||||||||| |||||||||||||||||| ||||||||||||||||||||||||||||
Db  11897  GAAGTCACAGGAAAAACACAAGCTGCGCAGAAGTGGTGCAAGAAACAGGATAAGTTACTC  11838

Qy   2325  GCCAAGTGCCTCAGGGCTATCAAGTTTCTGAAGGACCAAGATCAGCTGCTCCTCTTCCACT  2384
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db  11837  GTCAAGTGCCTCAGGGCTATCAAGTTTCTGAAGGACCAAGATCAGCTGCTCCTCTTCCACT  11778

Qy   2385  ACAACTATTCCGAACGACAGCTCCCTCAGGACATGCCTTCGAGAGATATGACGCGCCC  2444
           |||  | |||| | ||| ||| ||||||||||||||||||||||||||||||||||||
Db  11777  ACAGCCATTCCGAGGGACAACTTTCTCAGGACATGCCTTCGAGAGATATGACGCGCCC  11718

Qy   2445  ------------GAGCCTAGAGAGCAGAAGAATCTGCAAGAATTCCGCATGTCCCTGCGAGG  2483
                       |||||||||||||||||||||||||||||||||||||||||||||||||
Db  11717  GAGCCAAGTCGGCGGCAGGCCTGAGCCTAGAGAGCAGAAGAATCTGCAAGAATTCCGCATGTCCCTGCGAGG  11658

Qy   2484  CATTCATCATTCGAGCCTCGTGAATCTAGGAGGAACGACCACACTCACTCGATCT  2543
           ||||||||||||||||||||||||| -|||||||||||||||||||||||||||||
Db  11657  CATTCATCATTCGAGCCTCGTGAATCT-GGGAGGAGGAGGAGGACCACCACTCACTCGATCT  11599

Qy   2544  AGCAGCAGGAGGAGCAGGACACTTCTGCAGTTCTCGTAGTTCGTAGTTTACGCGGTGCTGCCGC  2603
           |||||||||||||||||||||||| || |||||||||| |||| ||||||||||||||||||||
Db  11539  AGCAGCAGGAGGAGCAGGACACTTCTACAGTTCTCGTAGTTTACGCCGTGCTGCCACCGC  11598

Qy   2604  AATAGAAGAAGAGAGGTGAGTATCCTCAGACGCGGTGCTGGCGCGCCACGAGCTGATGAG  2663
           ||| ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db  11479  AGTAGAAGAAGAGAGGTCGAGTATCCTCAGACGTATCCGCTGCCAGCGCGCCACGAGCTGATGAG  11538

Qy   2664  ATCGAGTACCCACATGCTGGAGCTGATACGGAACATGGCCGGGTTCGTCTATGGCTTGGGAG  2723
           ||||||||||||||||||||||||| |||||| ||||||||||||||||||||||||||||
Db  11478  GTCGAGTACCCACCCGCTGCTGGAGCTGATACGGAACAAGGCCGGGTTCGTCTATGGCCCTGGGAG  11419

Qy   2724  CAATCACAGCAGCAGCCATTGCAGCCAACTTGACTACCAACTTCGTTCATTCATTTGTCGACTGAGGTAAGCGCCTC  2783
           ||||||||||||||||||| ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db  11359  CAATCACACAGCAGCCATTGCAGCCAACTCCGTTCATTCTTCGACTAAGGTAAGCGCCTC  11418

Qy   2784  ACTTCACCATTATATATATCATCTCTGTGATTGTTCTTTATTTGTTCTTGTTCAGTGATTG  2843
           ||||||||||||||||||||||||| ||||||||||||||| ||||||| |||||||||||
Db  11358  ACTTCACCATTGTATATATCATCTGTTGTGATTGTTCTTCTTATTTGTGTATTTGTGTGATTG  11299

Qy   2844  GATTTGTCCCTGAGTACTCTCTCTCTCCCAAGTTTATTCACACAGTGGACTGTGCTGTGTGATTTAAGTT  2903
           ||||||||||||||||||||||||||||| |||||| ||||||||||||||||||||||||||||||||
Db  11239  GATTTGTCCTGAGTACTCTCTCTCCCAAGTTTATTCACACAGTGACTGTGCTGTGTGATTTAAGTT  11298

Qy   2904  TGGGGGAGGGCTCAGGAA--GTATGTTGCATTGTGTTGCATTTAGGATGTGTTCATTCATCATCT  2961
           |||||||||||||||||   ||||||| ||||||||||||| ||||||||||||||||||||||
Db  11238  TAGGGGAGGGCTCAGGAAGTGTGTTTGCATTGTGTTGCATTTAGGATGTGTTCATTCATCATCT  11179

Qy   2962  AAGGCATAGAAAAAAACCAAAAAAATTAAAAAATTTCAGAAAATGATTTCACAAAAAAAGA  3021
           ||||||||||||| ||||||||||||| ||||||||||| ||||||||||||||||||||
Db  11178  AAGGCATAGAAAAAACC-AAAAAATTTAAAAAAATTTAAGAAAATGATTTCACAAAAAAAGA  11120

Qy   3022  GTGTTCATGTAGTTGCATTCATTACATTGGATCAAGTCAAGTTGTTCATTTAGGATTGTT  3081
           |||||||||||||||||||||||||||||| || ||||||||||||||||||||||||||
Db  11119  GTGTTCATGTAGTTGCATTTGCATTCATTACATTTAGGATTCTCGAGTGTTCATTTAGGATTGTT  11060

Qy   3082  GCATATGCATAGGGGATAATGATCAGAGATAGCCTTGTAAGCA  3122
           ||||| ||||||||||||||||||||||||||||||||||||
Db  11059  GCATGTGCATAGGGGATAATGATCAGAGATAGCCTTGTAAGCA  11019
```

```
                DVDGNAPLHLATINWRCRTVDKLAAFASTETKILNIQNKDGLRPLDIAELNLQPDYVL
                RERLRTLMVLLCVYAPKSVGWLPTSGMTLRSRSEPLDAKKYKDHINALLLVATLVATVT
                FAAGTTIPGGFNSSAAPNMGMATLADDSTLFFFLVLDTLAMQSSIVAIVALI"
exon            complement(6741..7064)
                /gene="AT4g03480"
                /number=1
intron          complement(7065..7160)
                /gene="AT4g03480"
                /number=1
exon            complement(7161..7708)
                /gene="AT4g03480"
                /number=2
intron          complement(7709..7782)
                /gene="AT4g03480"
                /number=2
exon            complement(7783..8101)
                /gene="AT4g03480"
                /number=3
intron          complement(8102..8408)
                /gene="AT4g03480"
                /number=3
exon            complement(8409..8866)
                /gene="AT4g03480"
                /number=4
intron          complement(8867..9282)
                /gene="AT4g03480"
                /number=4
exon            complement(9283..9343)
                /gene="AT4g03480"
                /number=5
intron          complement(9344..9444)
                /gene="AT4g03480"
                /number=5
exon            complement(9445..9588)
                /gene="AT4g03480"
                /number=6
gene            10335..13501
                /gene="AT4g03490"
gene            complement(join(10335..10525,10983..11078,11590..12133,
                12345..12500,12627..12772,12871..13501))
                /gene="AT4g03490"
CDS             complement(join(10335..10525,10983..11078,11590..12133,
                12345..12500,12627..12772,12871..13501))
                /gene="AT4g03490"
                /note="contains similarity to mammalian ankyrins
                similarity to similar to A. thaliana hypothetical protein
                F21B7.8, GenBank accession number AC002560"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAB77834.1"
                /db_xref="GI:7270672"
                /db_xref="SPTREMBL:Q9ZT77"
                /translation="MGKVQEFEKVMEENEIPVLDQVTFQGNTIIHLAAIYGHDHLVRR
                ILAYELNIIRNWKRGLNCNFVPSFSHYQTLLVRRNYKGDLALHVAAAGHKLIVGLLI
                DCLRQLPQDITMVIGSEQMVIGNIFRVSNNDGNTALHLSLKGNHVSVSLQLVREDRST
                CFLLDKEDVSPLYMAAEAGYVSLVEHMLRGLDASFVGKSVLCAAVKSQNLDILTAVLE
                SDSDLVESRDEDGRTPLATAASIGYDIGVQHMLTRFASSTQGQNVLHVAAKSGNARAV
                GYLLRKSDVKRLINEQDIEGNTPLHLASSNSHPKVWLIWMALVAAGTTRAPRVHLRAD
                IPGLTTDEDLILIKIHKDRVNTLLVATLVATMAFAAGLSVPLGYNSTEFKSNVKHSYE
                ESAFHAFVICNSIAVYTAVISTVALIGTQLADLKCMLTTFKFTVPLLGFSIIAMSLAF
                VAGLYLVLGHHVWLAIFVLASGGFYLMALLLLIIPYASPYTFTLSRSLNSLVQNMSKE
                DVDSVNQLVPAPTEELALIKAIQVNRCIQGNIGVTLSLSYYLLHKHVAPINVLHLKLWN
                HYSLVFSGHLYDGPKKTYSQL"
exon            complement(10335..10525)
                /gene="AT4g03490"
                /number=1
intron          complement(10526..10982)
                /gene="AT4g03490"
                /number=1
exon            complement(10983..11078)
                /gene="AT4g03490"
                /number=2
intron          complement(11079..11589)

exon            complement(11590..12133)
                /gene="AT4g03490"
                /number=2
intron          complement(12134..12344)
                /gene="AT4g03490"
                /number=3
exon            complement(12345..12500)
                /gene="AT4g03490"
                /number=3
intron          complement(12501..12626)
                /gene="AT4g03490"
                /number=4
exon            complement(12627..12772)
                /gene="AT4g03490"
                /number=4
intron          complement(12773..12870)
                /gene="AT4g03490"
                /number=5
exon            complement(12871..13501)
                /gene="AT4g03490"
                /number=6
misc_feature    13514..15185
                /note="marker mi233"
gene            14170..17288
                /gene="AT4g03500"
CDS             join(14170..14204,15103..15848,15939..16477,16650..17288)
                /gene="AT4g03500"
                /note="hypothetical protein with some similarity to
                ankyrin
                similarity to"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAB77835.1"
                /db_xref="GI:7270673"
                /db_xref="SPTREMBL:Q9ZT78"
                /translation="MKNESRTSEEISKLGIFSGVGFSMGRPDVHLDRRVAENHQAIPR
                NNVRSTPSLDLSTLFDETSETKPMDPKTMAAVRAGKENYLRSNNSYISVAPTLVNDRG
                NTILHLAASSGHVSLVRYIIQKCPGLLLKSNMGEVALHLAAEAGHLDVVNNLLIDFIN
                DISCTNLPVAKRIYFAKNKNQDTALHVALKGKHEVVASYLVSAAKSLSFVANRDGFSP
                LYLAIEAGHTSLVTTMCHGTNELSSKVGGRSIVHAALKANRKDILDALLSKDASLINL
                RDEGRTSLSFGASIGYGYQGFSYLFDKNRDKVYVSDDDGLFPTHMAAKYGHVQILEEIL
                KHCPEAIELLDRDGQNILHLAAKYGKLKVIKFILSCCKDKNKKLNEQDVNGNTPLH
                LATINWHPKVVSMFTWDHRVDLKKRNYIGFTALDVAEENIDSSYIVHQRLFWMALINA
                GAPKSSTPITENLRSFKKPDGGKYKDRVNTLMLVATLVATMTFAGFTLPGGYNDSFP
                HLGMAVLAKRTAFQVFLVCDTLAMYSSIITIVALIWAQLGDLSIILKAFNIALPLGL
                ALTSMSIAFMAGTYVAVSHLPLLGYFVLGIGIFLLVLLLLVPSYAHAQPLLRH
                IFYYPYFLKLLAAGDNKNIVDVYTASDE"
exon            14170..14204
                /gene="AT4g03500"
                /number=1
intron          14205..15102
                /gene="AT4g03500"
                /number=1
exon            15103..15848
                /gene="AT4g03500"
                /number=1

Query Match     61.2%;   Score 1910.6;  DB 8;  Length 159629;
Best Local Similarity   90.1%;   Pred. No. 0;
Matches 2146;  Conservative 0;  Mismatches 174;  Indels 61;  Gaps 7;

Qy     796 GGTTACTCGACCGGTACGCTGGTTTAGATTGTGTTTTGTTTTGCTATTACTAACATA 855
           ||||||| ||||||||||||||||||||||||| ||||||||||||||  |||||||
Db  151121 GGTACACTCGACCGGTTACTCGTTGGATTGTGTTTTGTTTTTGTTGGAACTAACGTA 151062

Qy     856 TTAACGTTTATCTTGAGTTTCGTCGTTTTAGGTTTCATCATGAGTAACTACAGTGGA 915
           ||||||| |||||||| |||||| | |||||||||||||||||||||||||||||||
Db  151061 TTAACGTTAATTTTGAGTTACATTTGTTTTCAGGTTCATCATGAGTAACTACAGTGGA 151002

Qy     916 AAATCCTCTATGGACCCTGATTATAAATGTGGATGAAGCTAAGTCCTGGTCCACTAGACCG 975
           |||||||| ||||||||||||||||||| ||||||||||||| |||||||||||  |||
Db  151001 GAATCCTCTTCGGACCCTGATTATAAATGTGGATGAAGCTGAGTCCTGGTCCGTCTAGACC- 150943
```

Thu Sep 9 11:17:23 2004   us-10-088-384a-27.rge

```
Qy      976  GAGTGAGAGCAACATGTTTACGAGAGCTATAGGAGGATGAATTTGAACGCTTCTGCAGCTCGA  1035
             |||||||||||||||||||| |||||| ||| |||||||||||| |||||||||||||||||||
Db   150942  --GAGGAGCAACATGTTTACCAGAGCTTCAGGGATGAATTGAATGCTTCTGCAGCTCGA  150885

Qy     1036  CGTAATCAAGAAGAGCTGAAAATCGCCAGAGGCGATGCTGCAGTAGTCGAGTAGATATAGAG  1095
             |||||||| |||| |||||||||||||||||||||||   |||||||||||||||||||||||
Db   150884  CGTAATCAGAGAGTGCTGAAATCGCCAGAGGCGATGTCGAGTAGATATGAG  150825

Qy     1096  CTGATTGATGAGGATATCAAAACTGAGTATGAGCCAGAGTCATGGCGCAAGGAGGACGAAG  1155
             |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db   150824  CTGATTGATGAGGACATCGACGAACTGAGTATGAGCAAGAGTCATGGCGTAAGGACAACAAAG  150765

Qy     1156  CTACTGACAAAATCGACGAGGTTACAGTGGAGGTTATATCGAGATTCTTTGAGATGAAT  1215
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   150764  CTACTGACAAAATCGACGAGGTTACAGTGGAGGTTATATCGAGATTCTTTAAGGATGAAT  150705

Qy     1216  GACTTCTGGGGAACGAGGTATCCCTGAGATGAGACTTAGCCCAGTTGGGGTTACTGGAG  1275
             ||||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db   150704  GACTTCTGGGGAACAGGGTATTCGCTATAGAGACTTAGCCCAGTTGGGGTTACTAGAG  150645

Qy     1276  GAGCTGCAGCATCTGTTCGAGAGAAGTGTCATCTG------------  1308
             ||||||||||||||||||||||||||||||||||
Db   150644  GAGCTGCAGCATCTGTTCGAGAGAAGTGTCATCTGGAGAGACACTCATGTCTTACCCTTACGCT  150585

Qy     1309  ----ATAAGGAGGAGACAATCGAGTTTCTTCCACACTGCAAGTGGAAATGTATGAGGGA  1364
                 ||| |||||||||||| ||||||||||||||||||| ||||||||||||||||||
Db   150584  GCTTATTAGGAGGAGACAATCGAGTTTCTTCCACACTTGCAAGTGGAAATGTATGAGGGA  150525

Qy     1365  CTCACAGACTTGAGCTGGATACCATGGGCGTTAGGCTTCTTGACGTTCTTAGTGGTGATGAA  1424
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   150524  CTCACAGACTTGAGCTGGATACCATGGGCGTTAGGCTTCTTGACGTTCTTAGTGGTGATGAA  150465

Qy     1425  CAGCCGTACCAGTCAGCCGTATCAAGAGTTGGAGAAGAACTGTTGTTTCCCAGTGGAAAG  1484
             |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   150464  CAGCCGTATCAGTCAGCCGTATCAAGAGTTGGAGAAGAACTGTTTGGTTTCCCAGTGGAAAG  150405

Qy     1485  GGAACCAACCCCAGGTTTGACAGGAAGAGCTTAAGGATTTGTGGGGCTACTATTGGGAAC  1544
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   150404  GGAACCAACCCCAGGTTTGACAGGAAGAGCTTAAAGGATTTGTGGGGCCACTATAGGGAAC  150345

Qy     1545  AATCTTACCGGCTAAATCGACGCGGTCCAAGACCAACCAAATCCGGAGTCCTGTGATTCGC  1604
             |||||||||||||| |||||||||||||||||||||||||| |||||||||||||||| |||
Db   150344  AATCTTACCGGCTAAATCGACGCGGTCGCAAGACCAACCAAATCCGTGTGATTCGC  150285

Qy     1605  TACTTTCAGCGGCTCGGTTGCCAATGTTTTTACTCCAGGAGTTCACAGGCACCGTGTCT  1664
             |||||||||||||||||||||||||||||   |||||||||||||||||||||||||||||
Db   150284  TACTTTCAGCGGCTCGGTTGCCAAGTCTTTTTACTCCAGGAGTTCACAGGCACCGTGTCT  150225

Qy     1665  AACACAGACATGAAGATGATGATCAGGCGCTTATAGGGATTCTCCCGCCTTACAAAAGA  1724
             ||||||||||||||||||||||||| |||||||||||||||||||||||| |||||||||
Db   150224  AACACGGACATGGAGATGATGATCCGGCGCTTGTGGGAGCTTACGAAAGGA  150165

Qy     1725  AAGAAATGTCCTGAGAGGAGATCTTAACGACTCACCACCACCACCAGCCTCTGTTGATCCAT  1784
             |||||||||||||||||||||||||||||||||||| ||||||||||| ||||||||||||
Db   150164  AAGAAATGTTCTGAGAGGGAGATCTCAACACGACCACCGCCTCTGTTGATCCAC  150107

Qy     1785  CTGTGTGTGGGTACATGAAGTGGGCGCGTTAGACGGGCAGCTTCCGCTCAAGCACACTA  1844
             ||||||||||||||||||||||||||||| ||||||||||| |||||| |||||||| |||
Db   150106  CTGTGTGTCGGATATAGGAAGTGGGCGCGTTGATACGGCAGGTGAGAAGGTGAGAGGAGCACTA  150047

Qy     1845  TGCGTGGGTGGCCTTGTACGCGCGTTGCCAAGCGCCAATTCGAAAGTTCCGCTCAAGGACTA  1904
             |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   150046  TGCGTGGGTGGCCTTGTACGCGCGTTGCCAGCCAATTCGAAAGTTGTGGAGTTCCGCTCAAGGAGACCA  149987

Qy     1905  GGGTTAGCACCGAGAATGATGAGATGATGGATCACTTGGCCGATGTGAGTTCTTCAGTGTT  1964
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   149986  GGTTAGCACTGAGGATGGATGGATCAAGTGAGTTCTTGGAGTTT  149927

Qy     1965  GACATGGTTGCGACTTCACCCGCTACACCGCTCCACCGCTAGAATCGCCAAC  2024
             ||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   149926  GACATGGCTGCGACTTCACCCGCTACACCGCTCATCGATTCATCGATTAGAATCGCCAAC  149867

Qy     2025  ATTCTTTTCCCCTGCATTTACGCTACTAGGATTCTCGAGGGCCAGGAACATTGACTTCAAG  2084
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   149866  ATTCTTTCCCCTGCATTTATTGACGCTACTAGGATTCTCGAGGGGAAGGAACATTGACTTCAAG  149807

Qy     2085  CCTGCGCTTGAAGATCTTTATTCGAGGGCACTACTAGGGCTACTGAGTCAGATTAGTCACACC  2144
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   149806  CCTGCGCTTGAAGATCTTTATTTGAGGGCAGTACTAGGGGCTACTGATGTGTTAGTCACACC  149747

Qy     2145  GAAGGAGCTACAAATAGAAGAGTGTGATGAGACATATGATGATATAGATGAGGGGAGTTGAC  2204
             ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||| |||
Db   149746  GAAGGAGCTCCACGAGAGTGTGATGAGACAGATGATATAGATGAGGGGAGTTCAAC  149687

Qy     2205  ACGAGCATGTATCATTCAGTGAGCATATACCTCCAGCGAGGAAAAGCAAGAGTTGAGC  2264
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   149686  ACAAGCATGTATCATTCAGTGAGCATATACCTCCAGCGAGGAAAAGCAAGAGTTGAGC  149627

Qy     2265  GAAGCTCACAGGAACAACAGCAAGCTGCAGAAGTGGTGCAAGAAACAGGATAAGTTACTC  2324
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   149626  GAAGCTCACAGGAACAACACAGCAAGCTGCAGAAGTGGTGCAAGAAACAGGATAAGCTACTC  149567

Qy     2325  GCCAAGTGCCTCAGGGCTATCAGATCAAGTTTCTGAAGGACACATGCCTTCGAGGAGATCAGCGCGCCCC  2384
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   149566  GTCAAGTGCCTCAGGGCTATCAGTACAGTTTCTGAAGGACACATGCCTTCGAGGAGATCAGCGCGCCCC  149507

Qy     2385  ACAACTATTCGCAATGACAGTCTCCCTGAAATGTATGAGGGA  2444
             |||||||| ||||||||||||||||||||||||||||||| ||||
Db   149506  ACAGACCATTCCCAGGGACACAACTTCTCAGAGCCTCGTGAATCT--GGAGGAGGATGAGC  149447

Qy     2445  --------------------------GAGCCTAGAGACCTCGTGTAGTTCTGACGTTCTTAGTGGTGATGAA  2483
                                       ||||||||||||||| |||||||||||||||||||||||
Db   149446  GAGCCAAGTCGGCGCAGGCCTAGCCTCGTGAATCTTAGGAGGACTTAGGAGGACCTGTAGTTCGACGTCTCGT  149387

Qy     2484  CATTCATCATTCGAGTCTGTGAATCCGAGTCAGTCTCGCAGTTCTGCAGTTCTTAGTGAGTTCAGT  2543
             |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db   149386  CATTCATCATTCGAGTCTGTGAATCT-GGAGGAGGAGGAACAAGGACCACCACTGCTGCGCGC  149328

Qy     2544  AGCAGCAGGAGCAGACGACTTCAGTCGTGGGCTACTATTGGGAAC  2603
             |||||||||||||||||||||||||||||||||||||||||||||
Db   149327  AGCAAGAGGAGCAGACGACTTCTACAGTCGTTACGGCGGCGTGCTGCGCGC  149268

Qy     2604  AATAGAAGAGAGAGTCGAGTATCCTCAGACCGTGGCCGCGCCACAGAGCTGATGAG  2663
             |||||||||||| ||||||||||||||||||||||||||||||| |||||||||||
Db   149267  AGTAGAAGAAGAGGTCGAGTATCCTCAGACGGTGTCGGCGTCGCCAGAGCTGATGAGAG  149208

Qy     2664  ATCGAGTACCCACATGCTGCTGCTGGAGCTGATACGGAACAAGCGGCGGTTCGTCTATGGCTCTGGGAG  2723
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   149207  GTCGAGTACCCACCTGCTGGAGCTGATACGGAACAAGCGGCGGTTCGTCTATGGCCTCGGGAG  149148

Qy     2724  CAATCAGCAGCCATTGACTACCAACTTCGTTCATTCGACTGAGCGTAAGCGCCTC  2783
             |||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   149147  CAATCACATGCAGCCATTGACGACCAACTTCCGTTCATTCGACTAAGTGAAGCGCCTC  149088

Qy     2784  ACTTCACCATTATTATTATATCTTTGTGATTTGTTTCTTTATTTTGTTCAGTGATTG  2843
             |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db   149087  ACTTCACCATTGTTATTATCATCTGTGTGATTTGTTTCTTTATTTTGTTCTCTGTGATTG  149028

Qy     2844  GATTTGTCCTGAGTACTCCTTCCTCCAAGTTTATTCACAGTGACTGTGTGATTTAAGTT  2903
             ||||||||||||||||||||||| |||||||||||||||||| |||||||||||| |||||
Db   149027  GATTTGTCCTGTGTACTCGTTCCTCCCAAGTTTATTCACAGTGGTCTGTGTGATTTAAGTT  148968

Qy     2904  TGGGGAGGGCTCAGGAA--GTATGTTGCATTGTATATTTTAAGTCTGCATTCATCT  2961
             |||||||||||||| ||  |||||||||||||||||||| |||||||||||||||
Db   148967  TAGGGAGGGCTCAGGAAGTGTGTTGCATTGTATATATATATAAGAAAATGCTGCATTCATCT  148908

Qy     2962  AAGGCATAGAAAAGACCAAAAAATTAAAAATTCAGAAAATAAAAAGA  3021
             |||||||||||||||| ||||||| ||| ||||||||||| ||||||
Db   148907  AAGGCATAGAAAAAACC-AAAAATTTAAAAAAATTTAAGAAAAATTTAACAAAAAGA  148849

Qy     3022  GTGTTCATGTAGTTGCATTGTCATTACATTTGAGTCTAGGATCAAGCTGCCAAC  3081
             ||||||||||||||||||||||||||| ||||||||||| |||||||||||||
Db   148848  GTGTTCATGTAGTTGCATTACATTTGAGTCTAGGATTAGGATTAGGTTCTAGAGTTCAAGAGCCTTGTGAAGCA  148789

Qy     3082  GCATATGCATAGGGGATAATGATGAGATAGCCTTCAAG  3122
             |||||||||||||||||||||||||||||||||||||||
Qy
```

```
Db      148788  GCATGTGCATAGGGGATAATGATGAGATAGCCTTGTAAGCA 148748

RESULT 15
AC074111/c
LOCUS       AC074111              98536 bp    DNA     linear   PLN 16-AUG-2002
DEFINITION  Arabidopsis thaliana chromosome 1 BAC T18N24 genomic sequence,
            complete sequence.
ACCESSION   AC074111
VERSION     AC074111.2  GI:22267605
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 98536)
  AUTHORS   Town,C.D., Haas,B.J., Maiti,R., Hannick,L.I., Chan,A.P.,
            Ronning,C.M., Smith Jr.,R.K., Arbogast,T., Tallon,L.J.,
            Utterback,T.R., VanAken,S.E., Feldblyum,T.V., Yu,C., Wortman,J.R.,
            White,O. and Fraser,C.M.
  TITLE     Arabidopsis thaliana chromosome 1 BAC T18N24 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 98536)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
REFERENCE   3  (bases 1 to 98536)
  AUTHORS   Town,C.D. and White,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     On Aug 16, 2002 this sequence version replaced gi:9186844.
            Address all correspondence to:at@tigr.org

            BAC clone T18N24 is from Arabidopsis thaliana chromosome 1
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of several methods: Gene
            prediction programs including Genscan+ (Chris Burge,
            http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky,
            http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant
            of GlimmerM, see Mihaela Pertea,
            http://www.tigr.org/softlab/glimmerm/glimmerm.htm/glimmerm.html, and
            GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
            mpertea@tigr.org), searches of the complete sequence against a
            peptide database and the plant EST database at TIGR
            (http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
            indicate the level of evidence for their annotation. Genes with
            similarity to other proteins are named after the database hits.
            Genes without significant peptide similarity but with EST
            similarity are named as unknown proteins. Genes without protein
            or EST similarity, that are predicted by more than two gene
            prediction programs over most of their length are annotated as
            hypothetical proteins. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
            Simple repeats are identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES             Location/Qualifiers
     source          1..98536
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="1"
                     /map="NULL"
                     /clone="T18N24"
     gene            61..2215
                     /gene="T18N24.40"
                     /note="contains similarity to auxin independent growth
                     promoter (axi 1) GI:559921 from (Nicotiana tabacum)"
     mRNA            join(<61..115,191..297,385..473,553..750,836..1040,
```

```
     CDS             1144..1311,1390..1557,1639..1829,1984..>2215)
                     /gene="T18N24.40"
                     join(61..115,191..297,385..473,553..750,836..1040,
                     1144..1311,1390..1557,1639..1829,1984..2215)
                     /gene="T18N24.40"
                     /codon_start=1
                     /product="growth regulator-related protein"
                     /protein_id="AAM94943.1"
                     /db_xref="GI:22267609"
                     /translation="MPKPCPNSTVSSHFGLNRGINVFDEIFSESSALAPSSGYLIVEA
                     NGGLNQQRSAICNAVAVAGLLNAVLVIPRFEFHAIWKDSSNFGDIYDEDHFISSLEGY
                     VKIVRDVPDEIMTRFSYNVSSIPTIRVQAWATVNYYNGEVYPVLKEHGVIRITPFANR
                     LAMSVPPYIQLLRCIANYKALKFSSPISTLAEKLVDRMVEKSSATGGKYVSVHLRFEE
                     DMVAFSCCLYEGGRAEKSEMDVIRQKSWKGKFKRDRVIRPDLNRVNGKCPLTPLEVG
                     MMLRGMGFDNNTSIYLASGRIYQPEKHLAPLQEMFPRLYTKESLATPEELAFPQGYSS
                     RMAALDTVSLLSEVFVTTQGGNFPHFLMGHRRFLFGGHAKTVIPDKPKLVLLLQDME
                     MRWEVFKKEMKLMLGESDRKGVMVPRVRKINRKTSIYTYPLPECECIFHLSSNFSNTG
                     NILSLGALHPSSNLISSARL"
     repeat_region   complement(2876..2897)
                     /rpt_family="AT_rich"
     repeat_region   complement(3003..3070)
                     /rpt_family="(CAAAC)n"
     repeat_region   complement(4079..4172)
                     /rpt_family="AT_rich"
     gene            10087..14010
                     /gene="T18N24.39"
                     /note="pseudogene, hypothetical protein"
                     /pseudo
     repeat_region   complement(14695..14750)
                     /rpt_family="AT_rich"
     gene            15610..18756
                     /gene="T18N24.38"
                     /note="pseudogene, hypothetical protein"
                     /pseudo
     repeat_region   complement(16549..16704)
                     /rpt_family="(CATAA)n"
     gene            19928..25390
                     /gene="T18N24.37"
                     /note="pseudogene, putative Athila ORF 1"
                     /pseudo
     repeat_region   complement(25510..25533)
                     /rpt_family="(CA)n"
     gene            complement(28535..29290)
                     /gene="T18N24.44"
                     /note="contains similarity to transposon like ORF
                     GI:6069576 from (Brassica rapa)"
     mRNA            complement(<28535..>29290)
                     /gene="T18N24.44"
     CDS             complement(28535..29290)
                     /gene="T18N24.44"
                     /codon_start=1
                     /product="transposon-related protein"
                     /protein_id="AAM94941.1"
                     /db_xref="GI:22267607"
                     /translation="MTFANSHPIVQSSRGKFHCPCSVCKNEKHIISGRRVSSHLFSQG
                     FMPDYYVWYXHGEMNMDIRTSYTDRTYFSTNHEEVDNVVEDPYMDMVNDAFNFNVG
                     YDDNFHHDDSYONVEEPVRNHSNKFYDLLEGANNPLYDGCREGOSQLSLASRLMHNKA
                     EYNMSEKLVNSVCEMFTDFLPEGNQATTSHYQTEKLMRNLGLPYHTIDICQNNCMLFW
                     KEDEKEDQCRFCGAQRWKPKDDRRSQIIMMVGF"
     gene            31586..35050
                     /gene="T18N24.36"
                     /note="pseudogene, hypothetical protein"
                     /pseudo
     repeat_region   complement(39561..39594)
                     /rpt_family="AT_rich"
     gene            complement(40609..43080)
                     /gene="T18N24.43"
                     /note="pseudogene, putative Athila ORF 1"
                     /pseudo
     gene            complement(47106..48956)
                     /gene="T18N24.42"
                     /note="pseudogene, hypothetical protein"
                     /pseudo
```

```
gene            complement(49931..51034)
                /gene="T18N24.41"
                /note="contains similarity to hypothetical proteins from
                (Arabidopsis thaliana)"
mRNA            complement(<49931..>51034)
                /gene="T18N24.41"
CDS             complement(49931..51034)
                /gene="T18N24.41"
                /codon_start=1
                /product="conserved hypothetical protein"
                /protein_id="AAM94940.1"
                /db_xref="GI:22267606"
                /translation="MPVEQRMVVDFFGRSYSGKEFHQLNAAEEGSPWYADHVNYLEVD
                SCQRKGNISRRNEMPQNPILEVDIFPDVWGIDFMGPFPSSYGNKYILVAVDVYSNRVEA
                IASPTNDAKVMLKLFKIIFPRFGVLRVVISDGGKHFINKVFENLLKKYGVHGKVTTP
                YHPQTSEQVEISNREIKTILEKTVGTTRKDWSAKLDDALWDNRTAFKTPIGTTLFNLL
                YGKSCHLPVELEYKAMWAVKLLNFDIKTTEEKRLIQLSDLDEIRLEAYESSKIYKERT
                KLFHDKKIIKDFQVGDQVLLFNSLLKFPGKLKSRWSGPFCITEVRPYGAVTLAGKS
                GDFTVNGQRLKKYLADQILPDVTSVHLQELLDD"
gene            complement(54952..57564)
                /gene="T18N24.34"
                /note="pseudogene, putative Athila ORF 1"
                /pseudo
gene            complement(59357..65317)
                /gene="T18N24.33"
                /note="pseudogene, putative Athila ORF 1"
                /pseudo
gene            68409..74024
                /gene="T18N24.32"
                /note="pseudogene, putative Athila ORF 1"
                /pseudo
repeat_region   74899..74958
                /rpt_family="(GGAGAA)n"
gene            complement(78391..78922)
                /gene="T18N24.35"
                /note="contains similarity to hypothetical proteins from
                (Arabidopsis thaliana)"
mRNA            complement(join(<78391..78522,78603..78773,78851..>78922))
                /gene="T18N24.35"
CDS             complement(join(78391..78522,78603..78773,78851..78922))
                /gene="T18N24.35"
                /codon_start=1
                /product="conserved hypothetical protein"
                /protein_id="AAM94942.1"
                /db_xref="GI:22267608"
                /translation="MMAQRSLFHGDSSVDEKTVSEGNEELDRMEIRMRMRYGENRRRD
                KGVLIELDCNAKVVVATSHDPVTSEKLIFSCPYKISDGSGRGSGFKRWNTVALCDEFD
                MIKSEMIEMKKDLEAANICELS"
repeat_region   complement(80158..80181)
                /rpt_family="AT_rich"
repeat_region   complement(80451..80487)
                /rpt_family="AT_rich"
repeat_region   80618..80661
                /rpt_family="(GAAAA)n"
gene            complement(80634..83339)
                /gene="At1g37210"
                /note="synonym: T18N24.2; pseudogene, putative Athila
                retroelement ORF1 protein"
                /pseudo
```

```
Query Match      61.0%;  Score 1905.4;   DB 8;   Length 98536;
Best Local Similarity 82.3%;  Pred. No. 0;  Mismatches 426;  Indels 72;  Gaps 9;
Matches 2323;  Conservative 0;  Mismatches 426;  Indels 72;  Gaps 9;
Matches 1464
```

```
Db  39879  TCAAGGAAACAACTCGACCTCGCGCCATTTCACTCACTCGACTCTGCTC-AGCAAGCGCCT  39821
           ||||                                ||||    |||||| |||||||||||
Qy    539  CATCTTCATCATCTCCCTTACTCGACCACCGTGCGTCTCGCTCACCATTGCCATTTAAA    598

Db  39820  TCACCACTTCTTAGTCATAACCACTCGACCAAGGTGCCATCTCCCAAATCCATTCATCAC  39761
           ||||          |||||||||||||||||||||||||||||||||||||||||||||||
Qy    599  AGTCCACTCGATTGTCAAAGAGAAGAGAGTGAAGCTCACCACCGCCACTCAACCGCCGCGT  658

Db  39760  TCGACCATGCGCTCTTCACTTGATCGGCTCTACACCATCTCCACCACCCGGCTTCTTGTTC  39701
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    659  TTCCCCTACACATTCAAACACTCGACACTGCTACCACGTGCTACCATCTCCACCACCCGGCT  718

Db  39700  CAAACCGCTTCACTATCTTTCGCCCTCAAGGTCACTCGATCGCAAGCAAGCACACTCCATTCA  39641
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    719  ACCATACACTCGACCAACAACTCTCAAAGTAAAAAAAAGAA--AAAAAAGTCAAAA    776

Db  39640  CACAACCAAGGAAATCACTTGACCTCGAGTTCACCAAACTGGCAAAGAAAAAAATAAAAAA  39581
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    777  CCGACAGTTCACTCAACCGGTTACTCGACCGGTACGCTCGTTTAGATTGTGTTTTTGG  836

Db  39580  ATAAAAGAAAAATATTTTTGTACACTCGACCAGTTTCTCTGTTGGTTTGGTTTTTGG  39521
                          |||||||||||||||||||||||||||||||||||||||||||||||
Qy    837  TTTTGCTATTACTAACATATTAAGTTTATCTTGAGTTCGTCTCGTTCGTTGTTTTTCAT  896

Db  39461  TTTTGTTGGGACTAACGTATTTATCTTAAGTTATCTTAAGTTACATTGTTTTCAGGTTTCAT  39461
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    897  CATGAGTAACTACAGTGGAAAATCCTCTATGACCCTGATTATAATGTGGATGAAGCTAA  956

Db  39460  CATGAGTAACTATAGTGGAGATCCTCAATGGATCCTCAATGTGATCGTTGATGAGTGAAGCTGA  39401
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    957  GTCCTGGTCCACTAGAGCCGGAGTGAGCAACATGTTTACGAGGCTATAGGGATGAATT  1016

Db  39401  GTCTTGGTCCACTAGAGCAGGAGAGAAAAAACACGTTTACGAGAGTTTACGAGATGAATT  39341
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1017  TGAACGCTCGCAGCTCGACGTTAATCAAAGAAGAGCTGAAATCGCTAGAGAAGAGAGGGC  1076

Db  39340  TGAACGCTTCGACGACGAGCTGAAATCGCTAGAGAAGAGCTGAAATCGCCAGAGAAGAGAGAGC  39281
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1077  GATGT-CGAGTAGATATGAGCTGATTGATGAGGATATCAAAACTGAGTATGAGCCCAGT  1135

Db  39221  GATGTGCGAGTAGATATGAGCTGATTGATGAGGATCGATGAGGACATCGAAACTGAGTATGAGCCCAGT  39221
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1136  CATGGGCCAAGGAGACGAAGCTACTGAACAAATCCGACGAGGTTACAGTGGGAGGGTATA  1195

Db  39161  CATGGCCAAGGAGACAAAGCTGCTGAACAAACCCGACGAGGTTACAGTGGGAGGGTATA  39161
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1196  TCAGATTCTTTGAGATGAATGACTTCTGGGGAACGAGAGTATCCCGATATGAGACTTTAG  1255

Db  39101  TCCGATTCTTTGAGATGAATGACTTCTGGGGAACGAGAGTATCCCGCTATTAGACTTTAA  39101
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1256  CCCAGTTGGGGTTACTGGAGCGTGCAGCATCGTTCGAGAGTGTCATCTG-------  1308

Db   3904  CCTAGTTGGGGTTACTGGAGCGTGCAACATCGTTGCAACATCTGTTCGAGAGTGTCATCTCGAGACAC  39041
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1309  ----------------------ATAAGGAGGAGACAATCGAGTTTCTTTCCCACACTGC  1344

Db  38981  TCATGTCTTACCCTTACCGCTGCCGTGTAAGGAGGAGAAGACAATTGAATTTCTTTCCCACACTGC  38981
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1345  AAGTGGAAATGTATGAGGGACTTCAGAGACTTTGAGCTGGATACCATGGGGTTAGGCTTCT  1404

Db  38921  AAGTGGAGAGTGTATGAGGGACTTTGAGCTGGATACCATGGGGTTAGGTTCT  38861
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1405  TGACGTTCTTAGTGGATGAACAGCGGGTACCAGATTGATCAAGAAATTGGAAGAACTGT  1464

Db  38861  TGACGTTCTTCAGTGGCTGAACAGTGGTATCAGCTTTCGATTAAGAAATTGGAAGAACAGT  38801
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1465  TTGGTTTCCCAGTGGAAGGGAACCACCCCAGTTTGCAGGGAAGAGCTTAAGGATT  1524

Db  38801  TTGGTTTCCCAGTGGAAGGGAACCACCCCAGTTTGCAGGGAAGAGCTTAAGGAATT  38801
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1525  TGTGGGCTACTATTGGGAACAATCTACCGCTCAAACTCACATGTCAGTCACTCGATCTCTCCCTCAC  1584
```

```
Qy    359  CTCGACCGCTGAACACTTATCACCTTCAAGCTCCTATCTCTTCATCGTTCATCCTTCCAACACCG  418
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  39987  CTTCACCCGTCGAACACTTATCACCTCGCAAGCTCATCATCTCTTCATTCCACTCTCGAGACATCACT-  39929

Qy    419  CTGCTCTCATCCCCCACGCAAAGCTTGTCATCGTCATCCTTCCTCCACTCATCATCACCGTTCACTCGAT  478
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  38860  CTGCTCTCATCCCCCACGCAAAGCTTGTCATCGTCATCCTTCCTCCATCATCACCGTTCACTCGAT  38801

Qy    479  TCAGCAACCAAAACTCGACCTCGTCTCTCTTGCCACTCATAGTCACTCGATCTCTCCTCAC  538
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  39928  -------CGACCATAGCTCTTCTTGCCTTCGTCTTCTGCCATCATCACCGCTTCGTTCCTCCCTTCT  39880
```

```
Qy   2644 GCCGCCACAGAGCTGATGAGATCGAGTACCCACATGCTGAGCTGATACCGGAACATGGCCG 2703
          |||||||||||||||||||||  ||||| ||||||||||| |||||| ||||| ||||||
Db  37661 GCTGCCACAGAGCTGATGAGAGCTGAGTCAAGTACCCACTTGCTGAGCGATGATACAGAACAGGGCG 37602

Qy   2704 GTTCGTCTATGGCTTGGGAGCAATCACAGGCAGCAGCCATTGACTACCACTTCGTTCATTAT 2763
          | |||||||||||  || ||||||||||||||||||||| ||| |||| || ||||||||
Db  37601 GCTCGTCTATGGCCTGAGAGCAATCACAGGCAGCAGCCAATTCGGTTAATTCT 37542

Qy   2764 TCGACTGAGGTAAGCGCCTCACTTCACCACTCCTTGTGATTTGTTCT 2823
          ||||||| ||||||||||||||||||||||| |||||||||||||||
Db  37541 TCGACTAAGGTAAGCGCCTCACTTCACCACTCTTGTGATTTGTTCT 37482

Qy   2824 TTATTTTGTTTCAGTGATGGATTTGTCCTGAGTACTCTTCTTCCAAGTTTATTCACACAG 2883
          ||||||||||| | |||| |||||||||||    |||||||||||||||||||||||||| 
Db  37481 TTATTTTGTTTCTATGATTGGATTTGTCCTCAATACTCTTCTTCCAAGTTTATTCACACAA 37422

Qy   2884 TGGACTGTGTGATTAAGTTGGGGAGGGCTCAGGAA--GTATGTTGCATTGTATATAT 2941
          ||||||||||||||  ||||| ||||||||||| || ||||| ||||||||||||||
Db  37421 TGGACTGTGTGATTTAAGTTTGGGGAGGGCTCATGAGAGTGTTTGTTGCATTGTATATAT 37362

Qy   2942 TTTTAAGTCTGCATTCATCTAAGGCATAGAAAAACCAAAAAAATTAAAAATTCAGAA 3001
          | ||||||||| ||||||||||||||||| || ||||||||||||| |||||||||||
Db  37361 TCTTGAGTCTACATTCATCTAAGGCATAGTAGTAAAACAAAAAAATTGAAAAATTTCAGAA 37302

Qy   3002 AATGATTTCACAAAAAAGAGTGTTCATGTAGTTGCATTACATTTAGGATCAAGTCTAGA 3061
          |||||||||||||||| |||| ||||||||||| |||||||||||||||| |||||||||
Db  37301 AATGATTTCACAAAAACAGAGTTTCATGTAGTT-CATTACATTTAGGATCGAGTCTAGA 37243

Qy   3062 GTGTTTCATTTAGGATTGTTGCATATGCATAGGGGATAATGATAATGAGATAGCCTTGTAAGC 3121
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  37242 GTGTTTCATTTAGGATTGTTGCATATGCATAGGGGATAATGATAATGAGATAGCCTTGTAAGC 37183

Qy   3122 A 3122
          |
Db  37182 A 37182
```

Search completed: September 9, 2004, 07:15:41
Job time : 11679 secs

```
Db  38800 TGTGGGTACTATTGGGAACAATTACCGCTAAACTCGGCGCGTCCAAGAGCAACCAAA 38741

Qy   1585 TCCGGAGTCCTGTGATTCGCTACTTCAGCGCTTGCCAATGTTTTTACTTCCAGGG 1644
          |||||||||||  |||  |||||||  ||||||  ||| ||||||||||||||||
Db  38740 TCCGGAGTCCTGGATCCGGATCCGCTCGGTGCCAAGAGTTTTTACTTCCAGGG 38681

Qy   1645 AGTCTACAGGCACCGTGTCTAACACAGACATGAAGATGATAGATTCAGCGCGTTATAGGGA 1704
          ||||||||||||| ||||||||||| ||||||| ||||||||||| ||||| |||| |||
Db  38680 AGTCTACAGGCACTGTGTCTAACACGGACATGGAGATGATAGATTCCCGCGCTAACATGA 38621

Qy   1705 TTCTCCCGCCTACAAAAGGAAAAGAAATGTCCTGAGAGAGATCTTAACGACTCACCACCAG 1764
          |||||||  ||||||| |||||||| |||| ||||||||||||| |||||||||||||||
Db  38620 TTCTCCG--CTTACAAAGGAAAAGAATGTCTTGAGAGAGATCTTCACCACTCACCACCAG 38562

Qy   1765 TAATGCCTCGTTGATCCATCTGTGTGGTACATGAAGTGGGCGCTGACAAACGGCAAGA 1824
          |||||||||||| |||||| |||||| |||| |||||||||| ||||||||||||||||
Db  38561 TAATGCCTCGTTATCCACTTGTGTGATACAGGACGTGGGCCATTGACAAACGGCAAGA 38502

Qy   1825 AGAAGGTAAGAGAGCACTATGCGTGGGTGGCGTTGTGACGCCAATTCTGAAAGTTTGTG 1884
          ||||||| ||||||||||||||||||||||||||||||||||||||||| |||| ||||
Db  38501 AGAAGGTGAGAGAGCACTATGCGTGGGTGGCGTTGTGACGCCAATTCTGGAAGCTTGTG 38442

Qy   1885 GAGTTCCGCTCAAGAAGTAGGGTTAGCACGGAATGATGATGGACTTGATCACTTCGCGCC 1944
          ||||  ||||||||||||||||||  |||| ||  ||||||||||||||||||  |||||
Db  38441 GAGTACCGCTCAAGAAGTAGGGTTTAGCACCGAGATGATGGACTTGATCATTGGCGCGCC 38382

Qy   1945 GATGTGAGTTCTCGAGTTGACATGGTTGGGGACTTCACCGCTCACAGGTTCGAGCATT 2004
          |||||||||||| |||  ||||||||||| |||||| ||||||||||||||||||||||
Db  38381 GATGTGAGTTCTTAGACATGGTTGGCGACTTCCAGCGCTCACAGGTTCGAGCATT 38322

Qy   2005 CATCGATTAGAATCGCCAACATTCTTTTCCCCTGCCAATTACGCTACCTAGGAGTTCTGAGG 2064
          |||||||||||||||||||  ||||| | ||||||||||||||||| || ||| |||| ||||
Db  38321 CATCGATTAGAATCGCCAATAATTCTTCTCCCCTGCCAATTACGCTACCTAGGGAATTCTCGAGG 38262

Qy   2065 GCAGGAACAATGACTTCAAGCCTGCGCTTGAAGATCTTATTTCGAGGGGGCCAGTCCGCCAA 2124
           ||||||||||||||||||||||||||  ||||||||  |||||||||| | ||| |||||
Db  38261 GAAGGAACAATGACTTCAAGCCTGCGCGTTGAAGATCTCTATTTCGAGGGGCTGTCCACCAA 38202

Qy   2125 CTGAGGAGATTAGTCACACCGAAGAGCTACAAATAGAGAGATGTTGATGAGACATATGATA 2184
          |||||||||||||||||||||||||| |||||  ||||| |||||||||||||| ||| |
Db  38201 CTGAGGAGATTAGTCACACCGAAGAGAGTTACAACAGAAGAGATGTTGATGAGACAGATAAAA 38142

Qy   2185 TAGATGAGGCGGAGTTTGACAGGATTGCACAGCATGTATCATTTCAGTGAGCATATACCTCCAGCGA 2244
           ||||||||| ||| |||||| |||||||||||||||||||||||||||||| ||||||||||||||
Db  38141 TTGATGAGGCAGAGATTGACACGAGGATTGCACAGCATGTATCATTTCAGTGAGAATATACCTCCAGCGA 38082

Qy   2245 GGAAAAGCAAGAGTTTGAGCGGAAGCTCACAGGAGCAGGAACAACAACAAGCTGCAGAAGTGGTGCA 2304
           ||||||||||||||||||||||||||||||||||||||||||||||||||  | |||||||||||||
Db  38081 GGGAAAAGCAAGAGTTTGAGCGGAAGCTCACAGGAGCAGGAACAACAACAAGTTCCAGAAGTGGTGCA 38022

Qy   2305 AGAAACAGGATAAGTTACTCCGCCAAGTGCCTCCAAGTGCCTCCAAGTTCTGAAGGACAAGA 2364
          |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db  38021 AGAAACAAGATAAGTTACTCCGCCAAGTGCCTCCAAGTGCCTCCAAGTTCTGAAGGACTAGA 37962

Qy   2365 TCAGCTGCTCCTCTTCCACTACAACTATTCGGCACATCCGCTCCCTCCCTCAGCACATGCCTT 2424
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db  37961 TCAGCTGCTCCTCTTCCACTACAACTATTCGGCACATCCGCTCCCTCCCTCAGCACATGCCCT 37902

Qy   2425 CGAGGAGATATGAGGC--------------GCCCGAGCCTAGAGAGCAGAAGA 2463
          |||||||| |||||| |             ||||||||||||||||||| ||||
Db  37901 CGAGGAGAGATGAGACGGCGCCTGAGCCAAGTCGGCCGCCTGAGCCTAGAGAGCAGGAGAGAGA 37842

Qy   2464 TTCTGCATGTCCCTGCGAGGCATTCATCATTGCGAATCTGAAATCTAGGAAGAATAGGA 2523
          || ||||||||||||||||||||||||||| |||| ||| | ||||||||||||||||
Db  37841 TTCCGCATGTCCCTGCGAGGCATTCATCATTCGAATGTGGGAGGAAGAGAGA 37782

Qy   2524 GAACGACACTCACTCGATCTCAGCACGCAGGAGCAGACGACTTCTGCCAGTTCTCGTAGTTTAC 2583
           |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  37781 GGACCAGACTCACTCGATCTCAGCACGCAGGAGCAGACGACTTCTGCCAGTTCTCGTAGTTTAC 37722

Qy   2584 GCGACCGCCGGTGCTGCCGTGCTGGCCGGTGCTGGCCCGAGTATGTCGAGTATCCTCAGAGCGTGCTG 2643
          ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db  37721 GCGACCGCCGGTGCTGCCGTGCTGGCCGGTGCTGGCCCGAGTATCCTCAGAGCGTGCTACTG 37662
```

```
            GenCore version 5.1.6
      Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    September  9, 2004, 02:31:27 ; Search time 1106 Seconds
                                          (without alignments)
                                          11991.749 Million cell updates/sec

Title:     US-10-088-384A-27
Perfect score: 3122
Sequence:  3122
1 ttcatatattcgacctcttc........atgagatagcctgtaagca 3122

Scoring table: IDENTITY.NUC
Gapop 10.0 , Gapext 1.0

Searched:  3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters:     6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  N_Geneseq_29jan04:*
            1: geneseqn1980s:*
            2: geneseqn1990s:*
            3: geneseqn2000s:*
            4: geneseqn2001as:*
            5: geneseqn2001bs:*
            6: geneseqn2002s:*
            7: geneseqn2003as:*
            8: geneseqn2003bs:*
            9: geneseqn2003cs:*
           10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Query Length	DB	ID	Description
	1	3122	100.0	3122	4	AAF30459	Aaf30459 Transcrip
	2	1638.4	52.5	96988	3	AAF22290	Aaf22290 BAC conta
	3	1594.6	51.1	110000	3	AAF22305_09	Continuation (10 o
	4	1588.2	50.9	83390	3	AAF22283	Aaf22283 BAC conta
	5	1588.2	50.9	90336	3	AAF22289	Aaf22289 BAC conta
c	6	1415.4	45.3	129021	3	AAF22296	Aaf22296 BAC conta
	7	1398.6	44.8	4860	4	AAF30458	Aaf30458 Transcrip
	8	1366.8	43.8	67087	3	AAF22280	Aaf22280 BAC conta
	9	1339	42.9	64415	3	AAF22279	Aaf22279 BAC conta
c	10	1311	42.0	83390	3	AAF22283	Aaf22283 BAC conta
c	11	1311	42.0	90336	3	AAF22289	Aaf22289 BAC conta
c	12	1230.8	39.4	2512	4	AAF30450	Aaf30450 Transcrip
	13	1215.4	38.9	92584	3	AAF22288	Aaf22288 BAC conta
c	14	1174.8	37.6	96988	3	AAF22290	Aaf22290 BAC conta
	15	1127.2	36.1	1997	4	AAF30451	Aaf30451 Transcrip
	16	1108.6	35.5	134499	3	AAF22286	Aaf22286 BAC conta
c	17	1066.2	34.2	163319	3	AAF22306	Aaf22306 Arabidops
	18	1044.0	33.5	129021	3	AAF22296	Aaf22296 BAC conta
	19	1018.0	32.6	79122	3	AAF22294	Aaf22294 BAC conta
c	20	1002.6	32.1	59590	3	AAF22281	Aaf22281 BAC conta
c	21	988.2	31.7	86584	3	AAF22292	Aaf22292 BAC conta
c	22	981.8	31.4	82588	3	AAF22301	Aaf22301 BAC conta
c	23	955.6	30.6	110000	3	AAF22305_08	Continuation (9 of
	24	751.6	24.1	1857	3	AAZ35276	Aaz35276 Arabidops
	25	513.2	16.4	163319	3	AAF22306	Aaf22306 Arabidops
c	26	469	15.0	95223	3	AAF22282	Aaf22282 BAC conta
	27	467.8	15.0	109973	3	AAF22298	Aaf22298 BAC conta
	28	449	14.4	40349	3	AAF22278	Aaf22278 BAC conta
c	29	413.6	13.2	110000	3	AAF22303_0	Aaf22303 Arabidops
	30	409.4	13.1	82138	3	AAF22305_10	Continuation (11 o
	31	409.4	13.1	94895	3	AAF22302	Aaf22302 BAC conta
c	32	408.6	13.1	103929	3	AAF22287	Aaf22287 BAC conta
c	33	408.6	13.1	134499	3	AAF22286	Aaf22286 BAC conta
	34	358.6	11.5	95223	3	AAF22282	Aaf22282 BAC conta
	35	357.6	11.5	86584	3	AAF22292	Aaf22292 BAC conta
c	36	343.6	11.0	64415	3	AAF22279	Aaf22279 BAC conta
	37	217.4	7.0	80450	4	AAF22295	Aaf22295 BAC conta
	38	214	6.9	903	4	AAF30454	Aaf30454 Transcrip
c	39	180.2	5.8	40349	3	AAF22278	Aaf22278 BAC conta
	40	160.4	5.1	96583	3	AAF22297	Aaf22297 BAC conta
c	41	137	4.4	82138	3	AAF22305_10	Continuation (11 o
	42	137	4.4	82588	3	AAF22301	Aaf22301 BAC conta
	43	137	4.4	94895	3	AAF22302	Aaf22302 BAC conta
c	44	137	4.4	110000	3	AAF22305_09	Continuation (10 o
c	45	122.8	3.9	110000	3	AAF22303_0	Aaf22303 Arabidops

ALIGNMENTS

```
RESULT 1
AAF30459
ID   AAF30459 standard; DNA; 3122 BP.
XX
AC   AAF30459;
XX
DT   29-MAY-2001  (first entry)
DE   Transcriptionally silent information clone TSI-A genomic DNA.
XX
KW   TSI-A; transcriptionally silent information; gene silencing; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200120010-A2.
XX
PD   22-MAR-2001.
XX
PF   14-SEP-2000; 2000WO-EP008994.
XX
PR   16-SEP-1999;    99GB-00021964.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Steiner A, Mittelsten Scheid O, Paszkowski J;
XX
DR   WPI; 2001-244815/25.
XX
PT   Novel transcriptionally silenced genes obtained from Arabidopsis useful
PT   for selecting plants which compared to wild type plant are impaired in
PT   transcriptional gene silencing.
XX
PS   Claim 1; Page 34-35; 36pp; English.
XX
CC   This present sequence is that of a transcribed region of Arabidopsis
CC   thaliana genomic DNA identical to nucleotides 65081-68202 of BAC F7N22.
CC   The sequence was isolated from an Arabidopsis genomic DNA database on the
CC   basis of homology to Arabidopsis transcriptionally silent information
CC   (TSI) sequences. It is located 5' to the cDNA region given in AAF30456.
CC   Comparison of transcriptional gene expression between an Arabidopsis line
CC   carrying a silent transgene present in multiple copies and its mutant
CC   derivative mom1 impaired in silencing of the transgene revealed 2 cDNA
CC   clones which were expressed in the mutant plants, but not in the parental
CC   or wild-type plants. Both clones were derived from the same family of
CC   transcripts referred to as TSI. Genomic templates encoding TSI are
CC   repetitive elements with mainly pericentromeric location and conserved
```

```
CC    organisation among various ecotypes. Transcriptional silencing of genomic
CC    TSI templates is specifically released in the mutant. Transcription of
CC    TSI can be used as a marker to identify a nucleic acid encoding genetic
CC    a plant. The invention provides a nucleic acid encoding genetic
CC    information which is silenced in wild-type plants but whose expression is
CC    turned on in plants defective in transcriptional gene silencing, and has
CC    the formula: RA-RB-RC, where RA, RB, RC constitute component nucleotide
CC    sequences selected from the present sequence and those given in AAF30450-
CC    58. Also provided are methods of selecting a plant that is impaired in
CC    transcriptional gene silencing, and of producing DNA representing at
CC    least part of a gene necessary to maintain silencing of another gene in a
CC    cell or plant
XX
SQ    Sequence 3122 BP; 855 A; 767 C; 691 G; 809 T; 0 U; 0 Other;

Query Match             100.0%;  Score 3122;  DB 4;  Length 3122;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 3122;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy         1 TTCATATAATTCGACCTCTTCTTCATTCTTGCATCCAAAAGACACAACAAGCCGCCATC 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 TTCATATAATTCGACCTCTTCTTCATTCTTGCATCCAAAAGACACAACAAGCCGCCATC 60

Db        61 GCTTTCCCTCACAACTCTCACTCGACCACCGCCGCCCGCTCTCACTTACTCGGCTTCATCG 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy        61 GCTTTCCCTCACAACTCTCACTCGACCACCGCCGCCCGCTCTCACTTACTCGGCTTCATCG 120

Qy       121 CTCTCATCGCCATCTCAACATACTCGACCTCGGCGATATCACTCGAGCTCGCCGCTTCT 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 CTCTCATCGCCATCTCAACATACTCGACCTCGGCGATATCACTCGAGCTCGCCGCTTCT 180

Db       181 CACCGCCTCTCCATGTCACCGGCTCCCCTCTCCAAGGAAACAACTCGAGCTCTCC 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       181 CACCGCCTCTCCATGTCACCGGCTCCCCTCTCCAAGGAAACAACTCGAGCTCTCC 240

Qy       241 ATTTCACTCACTCGACCTCTACCACCAAGCCGGCTTCACCACACTCGACCTCTTAACCACT 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 ATTTCACTCACTCGACCTCTACCACCAAGCCGGCTTCACCACACTCGACCTCTTAACCACT 300

Db       301 CGACCACCTTCACCATCAACCAATCAAATGTTTCTCCTCCATCTTCATCGTTTCCAACACCGCT 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       301 CGACCACCTTCACCATCAACCAATCAAATGTTTCTCCTCCATCTTCATCGTTTCCAACACCGCT 360

Qy       361 CGACCGGTGACACTCGACCTATCACTTCAAGCTCTCCAGCTTCACCGTTCCAACACCGCT 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 CGACCGGTGACACTCGACCTATCACTTCAAGCTCTCCAGCTTCACCGTTCCAACACCGCT 420

Db       421 GCTCTCATCCCCACGAAAGCTTGTCATCACCTCACTCGATTC 480
             ||||||||||||||||||||||||||||||||||||||||||
Qy       421 GCTCTCATCCCCACGAAAGCTTGTCATCACCTCACTCGATTC 480

Qy       481 AGCAACCAAACTCGACCTCGTCTCTCTTGCCACTCAGTCACTCTCCTCCACCA 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 AGCAACCAAACTCGACCTCGTCTCTCTTGCCACTCAGTCACTCTCCTCCACCA 540

Db       541 TCTTCATCATCTCCCTTACTCGACCACGTGCGTCTCGCTCACCGTGCTCCATTGCCATTTAAAAG 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       541 TCTTCATCATCTCCCTTACTCGACCACGTGCGTCTCGCTCACCGTGCTCCATTGCCATTTAAAAG 600

Qy       601 CTCACTCGATTGTCAAAGAGAGAAGAAGCTGAGCTCAACCACCGCCACTCGACCGCGTT 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 CTCACTCGATTGTCAAAGAGAGAAGAAGCTGAGCTCAACCACCGCCACTCGACCGCGTT 660

Db       661 CCCTTCTACAACTCAACATTCAACATCTCCACCATCTCCACACCCGCCTCTTGTTCAC 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       661 CCCTTCTACAACTCAACATTCAACATCTCCACCATCTCCACACCCGCCTCTTGTTCAC 720

Qy       721 CATACACTCGACCAACAACTCTCAAAGTAAAAAAGAAAAAAAGTCAAAACCGA 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 CATACACTCGACCAACAACTCTCAAAGTAAAAAAGAAAAAAAGTCAAAACCGA 780

Db       781 CAGTTCACTCAACCGGTTTACTCGACCGTACCGTCGGTTAGATTGTGTGTTTTTGGTTTT 840
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       781 CAGTTCACTCAACCGGTTTACTCGACCGTACCGTCGGTTAGATTGTGTGTTTTTGGTTTT 840

Qy       841 GCTATTACTAACATATTAACGTTTATCTTGAGTTTCGTCTGTTTTAGGTTTCATCATG 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 GCTATTACTAACATATTAACGTTTATCTTGAGTTTCGTCTGTTTTAGGTTTCATCATG 900

Db       901 AGTAACTACAGTGGAAAATCCTCATGGACCCGATTATAATGTGGATGAAGCTAAGTCC 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       901 AGTAACTACAGTGGAAAATCCTCATGGACCCGATTATAATGTGGATGAAGCTAAGTCC 960

Qy       961 TGGTCCACTAGACCGGAGTGAGAGCAACATGTTTACGAGAGCTATAGGGATGAATTTGAA 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 TGGTCCACTAGACCGGAGTGAGAGCAACATGTTTACGAGAGCTATAGGGATGAATTTGAA 1020

Db      1021 CGCTCGCAGCTCGACGTAATCAAAGAAGAGCTGAAATCGCTAGAGAAGAGGGGCATG 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1021 CGCTCGCAGCTCGACGTAATCAAAGAAGAGCTGAAATCGCTAGAGAAGAGGGGCATG 1080

Qy      1081 TCGAGTAGATATGAGCTGATTGATGAGGATATCAAAACTGAGTATGAGCCAGAGTCATGG 1140
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 TCGAGTAGATATGAGCTGATTGATGAGGATATCAAAACTGAGTATGAGCCAGAGTCATGG 1140

Db      1141 CGCAAGGAGACGAAGCTACTGAACAAATCCGACGAGGTTACAGTGGAGTATATCAGA 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1141 CGCAAGGAGACGAAGCTACTGAACAAATCCGACGAGGTTACAGTGGAGTATATCAGA 1200

Qy      1201 TTCTTTGAGATGAATGACTTCTGGGAACGAGGTATCCTGATATAGACTTTAGCCCAG 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 TTCTTTGAGATGAATGACTTCTGGGAACGAGGTATCCTGATATAGACTTTAGCCCAG 1260

Db      1261 TTGGGGTTACTGGAGCGTGCAGCATCGTTCGAAGAGTCGTCATCTGATAAGGAGGAGA 1320
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1261 TTGGGGTTACTGGAGCGTGCAGCATCGTTCGAAGAGTCGTCATCTGATAAGGAGGAGA 1320

Qy      1321 CAATCGAGTTCTTCTTCCACACTCGAAGTGGAAATGTATGAGGGAGACTTTGAGC 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 CAATCGAGTTCTTCTTCCACACTCGAAGTGGAAATGTATGAGGGAGACTTTGAGC 1380

Db      1381 TGGATACCATGGGTTAGGCTTCTTAGTTGGATGAACAGCGGTACCAGATTT 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1381 TGGATACCATGGGTTAGGCTTCTTAGTTGGATGAACAGCGGTACCAGATTT 1440

Qy      1441 AGATCAAGAAATTGGAAGAACTGTTTGGTTTCCCTAGTACTATTGGAACAATCTACCGCTAAACT 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 AGATCAAGAAATTGGAAGAACTGTTTGGTTTCCCTAGTACTATTGGAACAATCTACCGCTAAACT 1500

Db      1501 TTGACAGGGAAGAGCTTAAGGATTTGTGGGCTACTATTGGAACAATCTACCGCTAAACT 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1501 TTGACAGGGAAGAGCTTAAGGATTTGTGGGCTACTATTGGAACAATCTACCGCTAAACT 1560

Qy      1561 CGACGCGGTCCAAGAGCAACCAAATCCGGAGTCCTGTGATTCGGCTCAGCGGCTCGG 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 CGACGCGGTCCAAGAGCAACCAAATCCGGAGTCCTGTGATTCGGCTCAGCGGCTCGG 1620

Db      1621 TTGCCAATGTTTTTACTCGACACAGAGCTCACAGCCACCGTGTCTAACAGACATGAAGA 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1621 TTGCCAATGTTTTTACTCGACACAGAGCTCACAGCCACCGTGTCTAACAGACATGAAGA 1680

Qy      1681 TGATAGATTCAGCGCCTTATAGGGATTCCTCCGCGCCTTACAAAAGGAAAGAATGTCCTGAGAG 1740
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1681 TGATAGATTCAGCGCCTTATAGGGATTCCTCCGCGCCTTACAAAAGGAAAGAATGTCCTGAGAG 1740

Db      1741 GAGATCTTAACGACTCACCACCACCAGTAATGCCTCTGTTGATCCATCTGTGTGGGTACATGA 1800
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1741 GAGATCTTAACGACTCACCACCACCAGTAATGCCTCTGTTGATCCATCTGTGTGGGTACATGA 1800

Qy      1801 AGTGGGCGCTGACAAACGGCAAGAGGAAGAAGTAAGAGAGAAGAGGACCACTATGCGTTGGCGGTTG 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1801 AGTGGGCGCTGACAAACGGCAAGAGGAAGAAGTAAGAGAGAAGAGGACCACTATGCGTTGGCGGTTG 1860

Db      1861 TGACGCCAATTCTGAAAGTTGTGAAGTTCCGCTCAAGGAGTAGGGTTAGCACCGAGAA 1920
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1861 TGACGCCAATTCTGAAAGTTGTGAAGTTCCGCTCAAGGAGTAGGGTTAGCACCGAGAA 1920
```

```
Qy  1921  TGATGACTTGGATCACTTGCGCCGATGTGACTTCTCTGAGTTGACATGGTTGGCGACT  1980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921  TGATGACTTGGATCACTTGCGCCGATGTGACTTCTCTGAGTTGACATGGTTGGCGACT  1980

Qy  1981  TTCACCGCTACAGGTTGAGCATTCATCGATTAGAATCGCCAACATTCTTTTCCCCTGCA  2040
          ||||||||||||||||  |||||||||||||||||||||||||||||||||||||||||
Db  1981  TTCACCGCTACAGGTTCGAGCATTCATCGATTAGAATCGCCAACATTCTTTTCCCCTGCA  2040

Qy  2041  TTTACGCTACTAGGATTCTCGAGGGCAGGAACATTCGACTTCAAGCCTGCGCTTGAAGATC  2100
          ||||||||||||||||||||||||||||||||||||    |||||||||||||||||||||
Db  2041  TTTACGCTACTAGGATTCTCGAGGGCAGGAACATTCAGCCTGCGCTTGAAGATC  2100

Qy  2101  TTTATTTCGAGGGCAGTCCGCGCCAACTGAGGAGTAGTCACACCGAAGGAGCTACAATAG  2160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101  TTTATTTCGAGGGCAGTCCGCGCCAACTGAGGAGTAGTCACACCGAAGGAGCTACAATAG  2160

Qy  2161  AAGATGTTGATGAGACATATGATATAGATGAGGCGGAGTTTGACACGAGCATGTATCATT  2220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161  AAGATGTTGATGAGACATATGATATAGATGAGGCGGAGTTTGACACGAGCATGTATCATT  2220

Qy  2221  TCAGTGAGCATATACCTCCAGCCGGAGAAAGCAAGAGTTTGAGCGAAGCTCACAGGAACA  2280
          ||||||||||||||||||||||||||||||  |||||||||||||||||||||||||||
Db  2221  TCAGTGAGCATATACCTCCAGCCGGAGAAAAGCAAGAGTTTGAGCGAAGCTCACAGGAACA  2280

Qy  2281  ACAGCAAGCTGCAGAAGTGGTGCAAGAAACAGGATAAGTACTCGGCCAAGTGCCTCAGGG  2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281  ACAGCAAGCTGCAGAAGTGGTGCAAGAAACAGGATAAGTACTCGGCCAAGTGCCTCAGGG  2340

Qy  2341  CTATCAAGTTTCTGAAGGACAAGATCAGCTGCTCCTCTTCCACTACAACTATTCCGCCAAT  2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341  CTATCAAGTTTCTGAAGGACAAGATCAGCTGCTCCTCTTCCACTACAACTATTCCGCCAAT  2400

Qy  2401  GACAGCTCCCTCAGGACATGCCTTGGAGGAGATATGACGCGCCCAGCCTTAGAGAGCAGA  2460
          |||||||||  |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401  GACAGCTCCCCAGGACATGCCTTGGAGGAGATATGACGCGCCCAGCCTTAGAGAGCAGA  2460

Qy  2461  AGATTCTGCATGTCCCTGCCGGAGGCATTCATCATTCGAGCCTCGTGAATCTAGGAAGAATA  2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461  AGATTCTGCATGTCCCTGCCGGAGGCATTCATCATTCGAGCCTCGTGAATCTAGGAAGAATA  2520

Qy  2521  GGAGAACGACACTCACTCGATCTAGCAGCAGCAGACACTTCTCGAGTCTCGTCGTAGTT  2580
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521  GGAGAACGACACTCACTCGATCTAGCAGCAGCAGACACTTCTCGAGTCTCGTCGTAGTT  2580

Qy  2581  TACGCGACCGCGGTGCTGGCCGCCAATAGAAGAGAGGTCGAGTATCCTCAGAGCGCGGTG  2640
          |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db  2581  TACGCGACCGCGGGTGCTGGCCGCCAATAGAAGAGAGGTCGAGTATCCTCAGAGCGCGGTG  2640

Qy  2641  CTGGCCGCCACAGAGCTGATGAGCAATCGGAGCTGGATACGGAACATG  2700
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641  CTGGCCGCCACAGAGCTGATGAGCAATCGGAGCTGGATACGGAACATG  2700

Qy  2701  GCGGTTCGTCTATGGCTTGGGAGCAATCACAGGCAGCCATTGACTACCACTTCGTTCAT  2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701  GCGGTTCGTCTATGGCTTGGGAGCAATCACAGGCAGCCATTGACTACCACTTCGTTCAT  2760

Qy  2761  TATTCGACTGAGGTAAGCCGCCTCACTTCACTTCACCATTATTATATCATCTCTTGTGATTTGT  2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2761  TATTCGACTGAGGTAAGCCGCCTCACTTCACTTCACCATTATTATATCATCTCTTGTGATTTGT  2820

Qy  2821  TCTTTTATTTTGTTTCAGTGATTGATTTGTCCTGAGTACTCTCTTCCAAGTTTATTCACA  2880
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2821  TCTTTTATTTTGTTTCAGTGATTGATTTGTCCTGAGTACTCTCTTCCAAGTTTATTCACA  2880

Qy  2881  CAGTTGGACTTGTGTGATTTAAGTTTGGGGAGGGGCTCAGGAAGTATGTTGCATTGTATATA  2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2881  CAGTTGGACTTGTGTGATTTAAGTTTGGGGAGGGGCTCAGGAAGTATGTTGCATTGTATATA  2940

Qy  2941  TTTTTAAGTCTGCATTCATCTAAGGCATAGAAAAAACCAAAAAAAAATTAAAAATTTCAGA  3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2941  TTTTTTAAGTCTGCATTCATCTAAGGCATAGAAAAAAACCAAAAAAAAATTAAAAATTTCAGA  3000
```

```
Qy  3001  AAATGATTTCACAAAAAAGAGTGTTCATGTAGTTGCATTACATTTAGGATCAAGTCTAG  3060
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3001  AAATGATTTCACAAAAAAGAGTGTTCATGTAGTTGCATTACATTTAGGATCAAGTCTAG  3060

Qy  3061  AGTGTTCATTTAGGATTGTTGCATATGCATAGGGATAAATGATGAGATAGCCTTGTAG  3120
          |||||||||||||||||||||||||||||||||||||| |||||||||||||||||| |
Db  3061  AGTGTTCATTTAGGATTGTTGCATATGCATAGGGATAATGATGAGATAGCCTTGTAAG  3120

Qy  3121  CA  3122
          ||
Db  3121  CA  3122
```

RESULT 2
AAF22290
ID   AAF22290 standard; DNA; 96988 BP.
XX
AC   AAF22290;
XX
DT   20-MAR-2001 (first entry)
XX
DE   BAC containing repeats from centromeres 1-4 #13.
XX
KW   Centromere; microsome; vector; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200055325-A2.
XX
PD   21-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-US007392.
XX
PR   18-MAR-1999;   99US-0125219P.
PR   01-APR-1999;   99US-0127409P.
PR   18-MAY-1999;   99US-0134770P.
PR   13-SEP-1999;   99US-0153584P.
PR   17-SEP-1999;   99US-0154603P.
PR   16-DEC-1999;   99US-0172493P.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Preuss D, Copenhaver G, Keith K;
XX
DR   WPI; 2000-587529/55.
XX
PT   Recombinant DNA construct comprising a plant centromere, useful for
PT   producing stably inherited microsomes which can serve as vectors for the
PT   construction of transgenic plant and animal cells.
XX
PS   Claim 102; Page 550-572; 1449pp; English.
XX
CC   The present invention relates to a recombinant DNA construct of a plant
CC   (Arabidopsis thaliana) centromere. The constructs are useful for
CC   producing stably inherited microsomes which can serve as vectors for the
CC   construction of transgenic plant and animal cells expressing selected
CC   proteins such as hormones, enzymes, interleukins, clotting factors,
CC   cytokines, antibodies, and growth factors
XX
SQ   Sequence 96988 BP; 29654 A; 18289 C; 18523 G; 30522 T; 0 U; 0 Other;

Query Match   52.5%; Score 1638.4; DB 3; Length 96988;
Best Local Similarity   89.6%; Pred. No. 0;
Matches 1874; Conservative 0; Mismatches 156; Indels 62; Gaps 8;

```
Qy   1090  TATGAGCTGATTGATGAGGATATCAAAACTGAGTATGGAGCCAGAGTCATGGCCAAGGAG  1149
           ||  |     |||| |||| |||  ||||||||||||||||||||||||||||||||||||
Db  54120  TAAAAACCTATTATCAAGGACATAGAAACTGAGTATGGAGCCAGAGTCATGGCCAAGGAG  54179

Qy   1150  ACGAAGCTACTGAACAAATCCGACGAGGTTACAGTGGAGGAGTATATCAGATTCTTTGAG  1209
           || |||||| |||||||| ||| ||||||||||||||||||||||||||||||||||||||
Db  54180  ACAAAGCTGATGAACAACAAACCCGACGAGGTTACAGTGGAGGAGTATATCAGATTCTTTGAG  54239
```

RESULT 3
AAF22305_09
Continuation (10 of 11) of AAF22305 from base 900001 (Arabidopsis thaliana chromosome
WP Sequence split into 11 fragments LOCUS AAF22305 Accession Aaf22305

Fragment Name	Begin	End
WP		
AAF22305_00	1	110000
AAF22305_01	100001	210000
AAF22305_02	200001	310000
AAF22305_03	300001	410000
AAF22305_04	400001	510000
AAF22305_05	500001	610000
AAF22305_06	600001	710000
AAF22305_07	700001	810000

```
WP   AAF22305_08    800001    910000
WP   AAF22305_09    900001   1010000
WP   AAF22305_10   1000001   1082138

Query Match          51.1%;   Score 1594.6;   DB 3;   Length 110000;
Best Local Similarity 76.5%;  Pred. No. 0;
Matches 2140;  Conservative 0;  Mismatches 554;  Indels 103;  Gaps 11;
```

```
Db  25482 CAAGGAGACAAAGCTGTGAACAAACCTGACGAGGTTACAGTGGAGGAGTATATCAGATT 25541

Qy   1203 CTTTGAGATGAATGACTTCTGGGAACAGAGTATCCGTGATATGAGACTTTAGCCCAGTT 1262
Db  25542 CTTTGAGCTGAACGACATTTTGGGAATGAGCTATCCCTGCTATGAGAATTTAGCCCAGCT 25601

Qy   1263 GGGGTTACTTGGAGGACGTGCAGCATCGTCTTCGAGAAGTGTCATCTG------ 1308
Db  25602 GGGGTTACTTGAGGACGTGCAGCATCGTCGTGAAGTGTCATCTGGAGACACTTCATGTC 25661

Qy   1309 ------------------ATAAGGAGGAGACAATCGAGTTTCTTCCACACTGCAAGTGA 1351
Db  25662 TTACCCATACGTCGCTACAAGGAGGAGACGATCAAGTTTCTTCCCACAATGCAAGTGGA 25721

Qy   1352 AATGTATGAGGGACTCACAGATTTGAGCTGGATACCATGGGGTTAGGCTTCTTGACGTT 1411
Db  25722 GATGTATCAGGGGTCTCCACAGAATTTGAGCTAGGAGCGACACCATGGGGATTCTTGACATT 25781

Qy   1412 CTTAGTGGATGAACAGCAGCGGTACCAGATTTAGATCGAAGAAATTGGAAGAACTGTTTGGTTT 1471
Db  25782 TTCAGTGGATGAGCGAGCAGCGCTATCAGTTGTCGATCAAGAAGAACTGTTTGGTTT 25841

Qy   1472 CCCTAGTGCGAAAGGGAACCAACCCCAGGTTTGACAGGGAAGAGACTTAAGGATTGTGGGC 1531
Db  25842 CCATAGTGGAAAGGGAACCAAGCCCAAGTTTGACAGG-AGACTTTAAGGATTTGTGGGC 25900

Qy   1532 TACTATTGGGAACAAATCTACCGTGACACTCACGCGGTCCAAGAGCAACCAAATCCGGAG 1591
Db  25901 TACTATTGGGAACAATGTACCCGCTAAGTACGGTCCAAGAGCAACCAAATCTGAG 25960

Qy   1592 TCCTGTGATTCGCTACTTTCAGCGGTTGCCAATGTTTTTACTCCAGGGAGTTCTCCG 1651
Db  25961 TCCATTGATTCGCTACTTTCAGTGCTCAAGTAGCCAATGTTCTTTACTTCCAGGGAGTTCTAC 26020

Qy   1652 AGGCACCGTGTCTAACACAGACATGAGCATGAGATGATAGATTCAGCGCTTATAGGGATTCTCCG 1711
Db  26021 AGACATCGTGTCTAACACAGACACATGAGCATGAGATGTGATAGATTTCGGGCACTTACAAGCATTCTCCA 26080

Qy   1712 CCTTACAAAGGAAAGAATGTCCTGAGGGAGATCTTAACGACTCCACCAGTAAAGCC 1771
Db  26081 CCGTACAAAGGGAAAGAAGAGGTCCTGAGGGAGGTGATCTCAATGACGCGCCACCAGTAATGCT 26140

Qy   1772 TCTGTTGATCCATCTGTGTGGGTACATGAAGTGGGCGCTGACAAACGGCAAGAAGAAGGT 1831
Db  26141 TCTGTTGATCCACCTGTGTGATACAGGAAAATGGGCGTTGACAAACTGCAAGAAGAAAGT 26200

Qy   1832 AAGAGGAGCACTATGCGTGGCGGTGCGTTGTGACGGCAATTCTGGAAGTTGTGGAGTTCC 1891
Db  26201 GAGAGGAACACTATGCGTAGGTGCGTTGTGACGCCGATTCTGGAAGCTGTGTGGAGTACC 26260

Qy   1892 GCTCAAGGAGTAGGGTTAGCACCGGAGAATCATGGGGACTTGCACTGCGCCGATGTGA 1951
Db  26261 GCTCAAGGAGAAACAGGGGTTAGAGCCGGAGAATGGAATGGATTTGGATCATTGACGCCGATGTGA 26320

Qy   1952 GTTCTCTGAGTTTGACATGGTTGGGCGACTTTCACCGGTTCGGAGCATTCATCGAT 2011
Db  26321 GTTCTTGGAGCATGACATGGTTGGGCAACTTCCACAGTTCAAGCATTCATCGAT 26380

Qy   2012 TAGAATCGCCAACATTCTTTTCGCCATTTACGCTTACTACTAGGAGGTCTCGAGGGCAGGAA 2071
Db  26381 CAGAATCGCCAACACATTTTCTCCCCTACATTGACGCTTACCAGGAGTTCTCGAGGGCAGGAA 26440

Qy   2072 CATTGACTTCAAGCCTGCGCTTGAGACATCTTTATTTCGAGGGCAGTCGGAGGGCGATGTC 2131
Db  26441 TATTGACTTTCTAGCCTCACGCTTGAGACATCTTTATTTCAAGGGCAGTCGCCGCCAACTGGCG 26500

Qy   2132 GATTAGTCACACCGAAGGAGGCTACAATAGAAGATGTTGATGAGACATATG 2191
Db  26501 GATCAGTGCACACGGAGGAGGCTACAATAGAAGAAGTTGATGAGACAGATGATGA 26560

Qy   2192 GGCGGAGTTTGACACGAGCATGTATCATTTCAGTGAGACATACCTCCAGCGAGGAAAAG 2251
```

```
Db  26561 AGCGGAGTTTGACACAGAGCATGTCTCATTCAGTGAGCATGTACCTCCAGCGCGAGGGAAAG 26620

Qy   2252 CAAGAGTTTGAGCGAAGCTCACAGGAACAACAGAAGCTGCAGAGAAACA 2311
                ||||||| |||||| |||||||||||||||| |||| ||||||| |||||
Db  26621 CAAGAGTTTGAGTGAAGCTCACAGGAACAACAACAAGTTGCAGAAGTGGTGTAAGAACA 26680

Qy   2312 GGATAAGTTACTCGCCAAGTGCCTCAGGCCTATCAAGTTTCTGAAGGACAAGATCAGCTG 2371
                ||||||||  ||||||||||||||||  |||| |||||||||||| | || |||||||||
Db  26681 TGATAAGTTACTTGCCAAGTGCCTCAAAGCTATCAAGTTTCTGAAGGACTAAGCTG 26740

Qy   2372 CTCCTCTTCCACTACAACTATTCCGACAGCTCCCTCAGGACATGCCTTCGAGGAG 2431
                ||||||| |||||||| ||||||  ||||| | |||||||||||||||||||||||
Db  26741 CTCCTCTACCACTACAGCTATTCCGCATGGACAGCTCCCTCAGGACACATGCCTTCGAGGAG 26800

Qy   2432 ATATGACGC---------------GCCCGAGGCCTAGAGAGCAGCAGAGATTCTGCA 2470
                |||||||                |||||||||||||||||||||||||||| ||||
Db  26801 ATATGATGCGGCTCGAGACAAGTCGGCGCAGGCCTCGTGAGACAAGATTATGCA 26860

Qy   2471 TGTCCCTGCGGAGGCATTCATCATTCGAGCCTCGTGAATCTAGGAAGAATAGAGAACGAC 2530
                |||||||||||||||||||||||||||||||||||   || |  ||||||||||||||||
Db  26861 TGTCCCTGCGAGGCATTCATCATTCGAGCCTCGTGAATTGGGAGAAAGAGGAGGACCAA 26920

Qy   2531 ACTCACTGGATCTAGCACGCAGGAGCAGCAGACGACTTCTGCAGTCTCGTAGTTTACGCGACCG 2590
                ||||||||||||| |||||||||||||||||||||| |||||||||||||||||||||||||
Db  26921 ACTTACTGGATCTAGCAGCAGCAGAGCAGCAGACACTTCTGCAATCTCGTAGTTTACGCGACCG 26980

Qy   2591 CGGTGCTGGCCGCAATAGAAGAAGAGAGAGTCGAGTATCCTCAGAGCGCGTGCTGGCCGCCCA 2650
                   ||||| |||||||||||||||||||||||| ||||||||||||||||
Db  26981 CGG--------TAGAAGAGAGAGGTCGAGTATCCTCA-AGCGGTGCTGGCCGCCA 27027

Qy   2651 CAGAGCTGATGAGATCGAGTCACCCACATGCTGGAGCTGATACGGAACATGGCGGTTCGTC 2710
                ||||||||||||||||||||||||| |||||||||||||||||||||| | ||||||||||
Db  27028 CAGAGCTGATGAGATCGAGTCACCCACATGCTAGAGCTAGAGCTAGAGTGAGAACAGGGGCGATTCTTC 27087

Qy   2711 TATGGCTTGGGAGCAATCACAGGCAGCCATTGACTACCAACTTCGTTCATTATTCGACTG 2770
                |||||||| |||||||||||| |||||||||||||||||||||| |||||| | ||||||
Db  27088 TATGGCCTCGGAGCAATCACATGCAACCATTGACGACCAACTTCGTTCATTCTTCGACTG 27147

Qy   2771 AGGTTAAGCGCCTCACTTCACCATTATATATATCATCATCTCTTGTGATTTGTTCTTTATTTT 2830
                ||||||||||||||||||||||||||| |||| ||||||||||||||||||| ||||||||
Db  27148 AGGTAAGCGCCCTCACTTCTCCATTGTATATCATCTCTTGTGATTTGTTCTTTATTTT 27207

Qy   2831 GTTTCAGTGATTGGATTTGTCCTGAGTACTCTCTCTCCAAGTTATTCACACAGTGGACTG 2890
                |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db  27208 GTTTCTGTGATTTGATTTGTCCTGAGTACTCTCTCTCCAAGTTAGTCACACAGTGGACTG 27267

Qy   2891 TGTGATTTAAGTTTGGGGGAGGGGCTCAGGAAGTATGT 2927
                ||||||||||||||||||||||||||||||||||||||
Db  27268 TGTGATTTAAGTTTGGGGGAGGGGCTCAGGAAGTGTGT 27304
```

RESULT 4

AAF22283

ID	AAF22283	standard; DNA; 83390 BP.
XX		
AC	AAF22283;	
XX		
DT	20-MAR-2001	(first entry)
XX		
DE	BAC containing repeats from centromeres 1-4 #6.	
XX		
KW	Centromere; microsome; vector; ds.	
XX		
OS	Arabidopsis thaliana.	
XX		
PN	WO200055325-A2.	
XX		
PD	21-SEP-2000.	
XX		
PF	17-MAR-2000; 2000WO-US007392.	
XX		
PR	18-MAR-1999; 99US-0125219P.	
PR	01-APR-1999; 99US-0127409P.	

XX
PT Recombinant DNA construct comprising a plant centromere, useful for
PT producing stably inherited microsomes which can serve as vectors for the
PT construction of transgenic plant and animal cells.
XX
PS Claim 102; Page 386-404; 1449pp; English.
XX
CC The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for
CC producing stably inherited microsomes which can serve as vectors for the
CC construction of transgenic plant and animal cells expressing selected
CC proteins such as hormones, enzymes, interleukins, clotting factors,
CC cytokines, antibodies, and growth factors
XX
SQ Sequence 83390 BP; 24664 A; 17305 C; 17224 G; 24140 T; 0 U; 57 Other;

```
Query Match             50.9%;  Score 1588.2;  DB 3;  Length 83390;
Best Local Similarity   77.8%;  Pred. No. 0;
Matches 2174;  Conservative  0;  Mismatches  263;  Indels  358;  Gaps  7;

Qy    662 CCTCTACACATTCAACACTCGACCACGGTGCTACCATCTCCACACCCGCTCTTGTTCACC 721
              |||||||||||||||||||||||||||||   ||| ||||||||||||||||||||||||
Db  38908 CCTCTTCAAGTCACCACTCGAC--CTACGCATCACTCTTACCACGAAACGAACTTCACT 38965

Qy    722 ATACACTCGACCAACAACTTCTCAAGTAAAAAAAAAGAAAAAAAAGTCAAAACCGAC 781
              ||||||||||||||  ||||||||||||||||||| ||||||||||||||||||||||
Db  38966 CGGTCGTACGAAATCACTTGATCAACGACCGACCGGTTTACTGTGTTGTGTTGTGTTTTG 39025

Qy    782 AGTTTCACTCAACCGGTTACTCGACCGGTTAGCTCGGTTGATTGTTTTTTGGTTTTG 841
              |||||||  ||||  ||||| ||||||||||||||  ||||||||||||||||||||
Db  39026 AGGTTTACTCGATTACACTCACTCGACCGGTTTACTGTGTTGTGTTGTGTTTTGGTTTTTG 39085

Qy    842 CTATTACTAACATATTAACGTTTATCTTTGAGTTTCGTCTGTTTTTACACGTCGGACTG 901
              |||||||||||||||||||||||||||||| |||||||| ||||||||||||||||||||
Db  39086 TTGAGACTAACTTATTAACGTTTATCTTTGAGTTTCAATTGTTGATCATGA 39145

Qy    902 GTAACTACAGTCGGAAAATCCTCTATGAGCCTGATTATAATGTGGATGAAGCTAAGTCCT 961
              ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db  39146 GTAACTACAGTCGCAGAATCCTCTATGAGTCCTGATTATAATGTGGATGAAGCTGAGTCCT 39205

Qy    962 GGTCCACTAGACCGGAGTGAGAGCAACATGTTTACGAGCTATAGGAGCTTCAGGAATTTGAAC 1021
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  39206 GGTCCTCTAGACCGGAGTGAGAGCAACATGTTTACGAGCTATAGGAGCTTCAGGAATTTCAAC 39265

Qy   1022 GCTCTGCAGCTCGACGTAATCAAAGAGCTGAAATCGCTGAAATCACCGAGAGGAGGGCGATGT 1081
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  39266 GCTCCTCGTCTGCGCGTAATCGAGCGTAATCAGAGAGGAGCGAAATCACCGAGAGGAGCGATGC 39325

Qy   1082 CGAGTAGATATGAGTGAGCTGATTGATGAGGATATCAAAACTGAGTATGAGCCAGAGTCATGGC 1141
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  39326 CCAGTAGATATGAGCTGATTGATGAGGACATCAAAACTGAGTATGAGCCAAAGTCATCATGT 39385

Qy   1142 GCAAGGAGAGCGAAGCTACTGAACAAATCCGACGAGTTACAGTGGAGTATATCAGAT 1201
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  39386 GCAAGGAGAGCGAAGCTACTGAACAAACCCGAATAGGTTACAGTGGAGTGTATATCAGAT 39445

Qy   1202 TCTTTGAGATGAATGACTTCTGGGGAACGGGTATCCCTGATAATGAACTTTAGCCCAGT 1261
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  39446 TCTTTGAGATGAATGATTTATGGGGAACGAGGTATCCCTACTATGAGACTTTAGCCCAGC 39505

Qy   1262 TGGGGTTACTGGAGGACGTGCAGCATCTGTTCGAGAAGTGTCATCTG---------- 1308
              |||||||||||||||||||||||||||||||||||||||||||||
Db  39506 TGGGGTTACTGGAGGACGTGCAGCATCTGTTCGAGAAGTGGCATCTGGAACACTCATGT 39565
```

```
Qy   1309 --------------ATAAGGAGGAGCAATGAGTTCTTCTTCCACACTGCAAGTGG 1350
               |||||||||||||||||||||| |||||||||
Db  39556 CTTACTCTTACGCTTCCTATAAGGAGGAGACAATCGAGTTTCTTTCCACACTGCAAGTGG 39625

Qy   1351 AAATGTATGAGGGACTCACAGACTTTGAGCTGG------------ 1383
          ||||||||||||||||| ||| |||
Db  39626 AGATGTATGAGGGACTCACATACATTGAGTTTAGGGTTTAGGGTTTAGGGTTT 39685

Qy   1384 -------------------- 1383
Db  39686 AGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAAGGTTTAGGG 39745

Qy   1384 -------------------- 1383
Db  39746 TTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTA 39805

Qy   1384 -------------------- 1383
Db  39806 GGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGT 39865

Qy   1384 -------------------- 1383
Db  39866 TTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAG 39925

Qy   1384 -----------ATACCCATGGGTTAGGCTTCTTCTTAGTCGATGAACAGCGGT 1431
                     |||||||||||||||||||||||||
Db  39926 GGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTGGT 39985

Qy   1432 ACCAGATTTAGATCAAGAAATTGGAAGAACTGTTTGTTTCCCTAGTGGAAAGGGAACCA 1491
          |||||||||||||| ||||| |||||||||| ||||| |||||||||||||||||
Db  39986 ATCAGCTTTCGATCAAGAAGTTGGAAGAAGTTGGAAGAACTGTTTGTTTCCCTAGTGGAAAGGGAACCA 40045

Qy   1492 ACCCCAGGTTTGACACAGGAAGAGCTTAAGGATTTGTGGGCTACTATTGGGAACAATCTAC 1551
          ||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db  40046 AAACCAGGTTTGACACAGGAAGAGCTTAAGGATTTGTGGGCTACTATTGGGAATAATTTAC 40105

Qy   1552 CGCTAAACTCGACGCGGTCCAAGAGCAACCAAATCCGGAGTCCTGTGATTCGCTACTTTC 1611
          |||||||||| |||| ||||| |||||||||||||||||||||||||||||||||
Db  40106 CGCTAAACTTGGCGCGGTCGATCTCAAACAACCAACCACCAATCCGGAGTCCTGTGATCCGCTACTTTC 40165

Qy   1612 AGCGCTCGGTTGCCAATGTGTTTTACTCCAGGGAGTCTACAGGCCACCGTGTCTAACACAG 1671
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db  40166 ACCGCTCGGTTGCCAACGTGTTTTTACTCTAGGGAGTCTACAGGCCACCGTGTCTAACACGG 40225

Qy   1672 ACATGAAGATGATAGATTCAGCGCTTATAGGGATTCTCCGCCTTACAAAAGGAAAGAAATG 1731
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db  40226 ACATGGAGATGATAGATTCCGCGCTTACAGGGATTCTCTGCTGTACAAAGGGAAAGAAATG 40285

Qy   1732 TCCTGAGAGGAGATCTTAACGACTCACCACCAGTAATGCCTCGTCTCGTTGATCCATCGTGTG 1791
          ||||||||||||| ||||| |||||||||||||||||||||||||||||||||
Db  40286 TCTTGAGAGGAGATCTTAACGACTCATGCGCCTTATGCTGTGTGTTCGTTGATTGGATTTGT 40345

Qy   1792 GGTACATGAAGTGGCGGCTCACAAACGGCTCGACAAAGGGTAAGAGGAGCACTATGCGTGG 1851
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  40346 GATATAGGAAGTGGCGGCGTTGACGCCCAATTCTCGGAGGTTCGTGGAGTACGACGAGCACTATGCATGG 40405

Qy   1852 GTGGCGTTGTGACGGCCAATTCGAAAGTTTGTGGAGTTCCGCTCAAGGAAGTAGGGTTAG 1911
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  40406 ATGGCGTTGTGACGGCCAATTCTGGAGCTGGATCTCGGAGTCTCTGAGTGTCGTGG 40465

Qy   1912 CACCGAGAATGATGAGTCTGGATCCGTACACTTGGAGTCTCTGAGTTTGACATGG 1971
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  40466 CACCGAGAATGATGAGTCTGGATCACTTGGATCTGGAGTCTCTGAGTTTGACATGG 40525

Qy   1972 TTGGCGACTTCACCGTACGCTACTAGCTTCGAGATTCATCGATTAGAATCGGCCAACATTCTTT 2031
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  40526 CTAGCGACTTCACCGTACGCTACTAGGTTCGAGATTCATCGATTAGAATCGGCCAACATTCTTC 40585

Qy   2032 TCCCCTGAGTTACGCTACTAGGATTCTCGAGAGGGCAGCAGGAACCATTGACTTCAAGCCTGCGCC 2091
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  40586 TCCCCTACATTGATGCTACTAGGATTCTTGAGGGAGGGCAGCAGGAACCATTGACTTCAAGCCTGCGCC 40645
```

```
Qy   2092 TTGAAGATCTTTATTCGAGGGCAGTCCGCCAACTGAGGAGATTAGTCACACCGGAAGGAG 2151
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  40646 TTGAAGATCTTTATTCGAGGGCAGTCCGCCAACTGAGGAGATTAGTCACACCGGAAGGAG 40705

Qy   2152 CTACAATAGAAGATGTTGATGAGACATATGATATAGATGAGGCGGAGTTTGACACGAGCA 2211
          |||| || |||||||||||||||||||| ||||| ||||| |||||||||||||||||
Db  40706 CTATAACAGAGAGATGTTGATGAGACAGATGACATAGATAGATTTGACACGAGCA 40765

Qy   2212 TGTATCATTTCAGTGAGCATATACCTCCAGCGAGGGAAAAGCAAGAGTTTGAGCGAAGCTC 2271
          ||||||||||||||||||||||| |||| ||||||||||||||||||||||||||
Db  40766 TGTATCATTTCAGTGAGCATATACTTCCAGCGAGGGAAAGCAAGAGTTTGAGCGAAGCTC 40825

Qy   2272 ACAGGAACAACAGCAAGCTGCAGAAGTGTGCAAGAAACAGGATAAGTTACTCGGCCAAGT 2331
          |||||||||| ||||||||||||| ||||||||||||||||||||||||||||||||
Db  40826 ACAGGAATAACATCAAGCTGCAGAAGTGGTGCAAGAAACAGGATAAGTTTCTCGGCCAAGT 40885

Qy   2332 GCCTCAGGGCTATCAAGTTTCTGAAGGACAAGATCAGCTGCTCCTTCCACTACAACTA 2391
          |||||||||||||||||||| |||||||||||||||||||||| |||||||||| ||
Db  40886 GCCTCAGGGCTATCAAGTTTCTGAAGGATAAGATCAGCTGCTCCACTTCCACTACAGCTA 40945

Qy   2392 TTCCGCAATGACAGCTCCCTCAGGACATGCCTTCGAGAGATATGACCGCGCCC------- 2444
          ||||| |||||||||||||| |||||||||||||||||||||||
Db  40946 TTCCACAGGGACAGCTCCCTCAGGACATGCCTTCGAGAGAGATGACGCGCCCGAGCCAA 41005

Qy   2445 --------GAGCTAGAGAGCAGAGAATTCTGCATGTCCCTGCGAGGCATTCAT 2490
                  |||||||||||| |||||||||||||||||||||||||||||
Db  41006 TTCGGGCGCAAACCTGAGCCTAGGAGTTCTGCATGTCCCTGCGAGGCATTCAT 41065

Qy   2491 CATTCGAGCCTGGTGAATCTAGGAAGTCTAGGAGAAGAATAGAGAAGACACTCGATCTAGCAGCA 2550
          ||||||||||||| |||||||||||||||||| |||||| |||||||| |||||||||||
Db  41066 CATTCGAGCCTCATGAAATCTGGGAGACGTGATAATCGGGAGGAAAGAGAAGGAGCAATCAACA 41125

Qy   2551 GGAGCAGCAGACTTCTGCAGTCTCGTAGTTTACGGCACCGCGGTCGGCCGCAATAGAA 2610
          | || |||| ||| ||||||||||| |||||||||
Db  41126 AGGGTAGACGACTTCTGCATTCTCGTAGTTTAC-------------ATGAA 41163

Qy   2611 GAAGAGAGGTCGAGTATCCTCCAGAGAGCGGTGCCGCCACAGAGCTGTGAGATCGCGAGT 2670
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  41164 GAAGAGAGGTCGAGTATCCTCCAGAGCGTCGCGATGGCCGCCACAGAGCTGTGAGGTCGAGT 41223

Qy   2671 ACCCACATGCTGGAGTGATACGGATGGAACATGCGCGGTTCGTCGATCGCATCAGCAATCAC 2730
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db  41224 ACCCACCTACTGGAGTGCTGATACGGATGGAACAGGGCGGGTTCGTCTATGGCCCAGGGAGCAATCAC 41283

Qy   2731 AGGCAGCCATTGACTACCAACTTCGTTCATTATTCGTTCATTATTGTTCTTATTTGTTGT 2790
          ||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db  41284 AAGCAGCCATTGACGACCGCACTTCGTTCATTCATTATTGTTCTTATTCTTCACTTCAC 41343

Qy   2791 CATTATTATTATTATCATCTCTGTGATTTGTTCTTATTTGTTCTTCAGTGATTGATTTGT 2850
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  41344 CATTGTATTATATCATCTCTGTGATTTGATTTGTTCTTATTTGTTCTTCGAATGGATTTGT 41403

Qy   2851 CCTGAGTACTCTTCTTCCAAGTTAATTCACACAGTGCGACTGTGTGATTTAAGTTGTGGGGA 2910
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db  41404 CCTGGGTACTCTTCTTCCAAGTTATCCAAGTTATCCGACTGTGTGATTTTGTGATTTTGGGGA 41463

Qy   2911 GGGCTCAGGAAGAAGTAGTGTG---TTGCATTGTATATATTTTAAGTCATCATTCATTCAAGGCAT 2968
          ||||||||||||||||| ||||||||| |||||||||||||||||||||
Db  41464 GGGCTCAGGAAGAAGTGTGTGTATTGCATTGCATTGTATATTTTTGTATATTTTCACATAAGGCAT 41523

Qy   2969 AGAAAAACCAAAAAAAATTAAAAATTCAGAAAAATGATTTCAC-AAAAAAGAGTGTTC 3027
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db  41524 AGAAAAACCAAAAAAAAATTTTGAAAAATTTCAGAAAATTTCACAAAAATAAAGAATGATTT 41583

Qy   3028 ATGTAGTTGCATTACATTTAGGATCAAGTCTAGAGTGTTTCATTTAGGATTGTTGCATAT 3087
          |||||| |||||||||||||||||||||||| |||||||||||||||||||||
Db  41584 ATGTAGTTGCATTACATTTAGGATCGAGTCTAGAGTGTTTCATTTAGGATTTTGCATAT 41643

Qy   3088 GCATAGGGAATAATGAGTATG------------------------------- 3122
          ||||||||||||||||||||||||||||
Db  41644 GCATAGGGGATAATGATGAGATAGCCTTGTAAGCA 41678
```

```
RESULT 5
AAF22289
ID   AAF22289 standard; DNA; 90336 BP.
XX
AC   AAF22289;
XX
DT   20-MAR-2001 (first entry)
XX
DE   BAC containing repeats from centromeres 1-4 #12.
XX
KW   Centromere; michrosome; vector; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200055325-A2.
XX
PD   21-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-US007392.
XX
PR   18-MAR-1999; 99US-0125219P.
PR   01-APR-1999; 99US-0127409P.
PR   18-MAY-1999; 99US-0134770P.
PR   13-SEP-1999; 99US-0153584P.
PR   17-SEP-1999; 99US-0154603P.
PR   16-DEC-1999; 99US-0172493P.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Preuss D, Copenhaver G, Keith K;
XX
DR   WPI; 2000-587529/55.
XX
PT   Recombinant DNA construct comprising a plant centromere, useful for
PT   producing stably inherited michrosomes which can serve as vectors for the
PT   construction of transgenic plant and animal cells.
XX
PS   Claim 102; Page 529-549; 1449pp; English.
XX
CC   The present invention relates to a recombinant DNA construct of a plant
CC   (Arabidopsis thaliana) centromere. The constructs are useful as vectors for
CC   producing stably inherited michrosomes which can serve as vectors for the
CC   construction of transgenic plant and animal cells expressing selected
CC   proteins such as hormones, enzymes, interleukins, clotting factors,
CC   cytokines, antibodies, and growth factors
XX
SQ   Sequence 90336 BP; 26524 A; 18837 C; 18625 G; 26350 T; 0 U; 0 Other;

Query Match            50.9%;  Score 1588.2;  DB 3;  Length 90336;
Best Local Similarity  77.8%;  Pred. No. 0;
Matches 2174;  Conservative    0;  Mismatches  263;  Indels  358;  Gaps  7;

Qy      662 CCTCTACACATTCAACACTCGACCACGGTGCTACCATCTCCACACCCGCTCTTGTCACC 721
                |||||||||||||||||||||||||  |||||||||||||||||||||||||||||||
Db    51662 CCTCTTCAAGTCACCACTCGACC--CTACGCATCACTCTTACCACGAAACGAACTTCACT 51719

Qy      722 ATACACTCGACCAACAACTCTCAAGTAAAAAAAAGAAAAAAAAGAAAAAAAATTAAACGA 781
                |||||||||||||||||||||     |||        |||||||||||||||||
Db    51720 CGGTCGTACGGAACTCACCGGTTACTCACCACTCACTCGCGAAAGAGAAAAATTAAACGA 51779

Qy      782 AGTTTCACTCAACTCGATTCACTCATCATCACTCGACCGGTTACTGACCGGTTAGTGTTTG 841
                ||||||||||||  ||||||||||||||||||
Db    51780 AGGTTTACTCAGTTACTCACTCATCATCACTCGACCGGTCGTTTGTGTTTGTTGGTTTTG 51839

Qy      842 CTATTACTAACATATTAACGTTTATCTTGAGTTTCGTCTGTTTTTAGGTTTCATCATGA 901
                |||||||||||||||||||||||||||  |||||||    |||||||  |||||||
Db    51840 TTGAGACTTAACTTATTAACGTTTATCTTTGAGTTTCAATTGTTTACAGGTTTGATCATGA 51899

Qy      902 GTAACTACAGTCGAAATCCTCTATGGACCCTGATTATAATCGTAATATAAATGTGGATGA 961
                ||||||||||||| |||||||||||  |||||||||||||            |||||
Db    51900 GTAACTACAGTCGAGATCCTCTATGATCCTGATTATAATGTGGATGAGAGCTATAGGAGGA 51959

Qy      962 GGTCCACTAGACCGGAGTGAGAGCCAACATGTTTACGAGAGCTATAGGGATGAATTCGAAC 1021
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    51960 GGTCCTGCAGTCGACCGGAGAGAGAGAGAGAGACATGTTTACGAGAGCTTCAGGGATGAATTCAAC 52019

Qy     1022 GCTCTGCAGCTCGACGTAATCAAGAAGAGCTGAAATCGCTAGAGAAGGAAGAGGGCGATGT 1081
                ||||| ||||||||||||||||| |||||||||||||  |||||||||||||||||||||
Db    52020 GCTCCTCTGCTGACGTAATCAGAGAAGAGCAGAAATCACCAGAGGGAAGAGACGAGTCATGC 52079

Qy     1082 CGAGTAGATATGAGCTGATTGATTGAGTATCAAAACTGAGTATGCATCGAGTCATGGC 1141
                ||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||
Db    52080 CCAGTAGATATGAGCTGATTGATGAGCACATCAAAACTGAGTATGCATCGAGTCATGGT 52139

Qy     1142 GCAAGGAGGAGCAGAAGCTACTGAACAAATCCGACGAGGTTACAGTGGAGTATATCAGAT 1201
                ||||||| ||||||||||||||||||||| |||||| ||||||||||||||||||||||
Db    52140 GCAAGGAAGAGCAGAAGCTACTGAACAAACCCGAATAGGTTACAGTGGAGTATATCAGAT 52199

Qy     1202 TCTTTGAGATGAATGAGCTCTCGGGAACGAGGTATCCCTGATATGAGACTTTAGCCCAGT 1261
                |||||||||||||||||    |||||||||||||||||||| ||||||||||||||||
Db    52200 TCTTTGAGATGAATGAGATTATGGGAACGAGGTATCCCTACTATGAGACTTTAGCCCAGC 52259

Qy     1262 TGGGGTTACTGGAGGACGTGCAGCATCGTGTTCGAGAGTGTCATCTG--------- 1308
                |||||||||||||||||||||||||||||||||||||||| |||||
Db    52260 TGGGGTTACTGGAGGACGTGCAGCATCGTGTTCGAGAGTGGCATCTGGAAAACACTCATGT 52319

Qy     1309 ----------ATAAGGAGGAGACAATCGAGTTTCTTTCCACACTGCAAGTGG 1350
                          ||||||||||||||||||||||||||||||||||||||||
Db    52320 CTTACTCTTACGCTTCCTATAAGGAGGAGACAATCGAGTTTCTTTCCACACTGCAAGTGG 52379

Qy     1351 AAATGTATGAGGGACTCACAGACTTTGAGCTGG--------- 1383
                ||| |||||||||||||||| | |||||| | |
Db    52380 AGATGTATGAGGGACTCACATCATTGAGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTT 52439

Qy     1383 --------- 1383
Db    52440 AGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAAGGGTTTAGGG 52499

Qy     1383 --------- 1383
Db    52500 TTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTA 52559

Qy     1383 --------- 1383
Db    52560 GGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGT 52619

Qy     1383 --------- 1383
Db    52620 TTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTAG 52679

Qy     1384 ----------ATACCATGGGGTTAGGGTTAGGGTTCTTAGTGGATGAACAGCGT 1431
                          |||  || || ||| ||||||| ||||||||||
Db    52680 GGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTCTTCCAATGGATGATGGAGCAGTGT 52739

Qy     1432 ACCAGATTAGATCAAGAAATTGGAAGAACTGTTTGGTTCCTAGTGGAAGAGGGAACCA 1491
                ||| ||||| ||||||||||||||||||| |||||  ||||||||||||| ||||||||
Db    52740 ATCAGCTTCGATCAAGAGTTCGAAGAACTGTTTGTTTGTTTCCCTAGTGGAAGAGGGAACCA 52799

Qy     1492 ACCCCAGTTTGACAGGGAAGAGCTTAAGAGATTGTGGGCTACTATTGGGAACAATCTAC 1551
                ||  |||||||||||||||||||||||||||||||||||||||||||||||||  ||||
Db    52800 AAACCAGTTTGACAGGGAAGAGCTTAAGGATTTGTGGGCTACTATTGGGAATAATTTAC 52859

Qy     1552 CGGTCAAACTCGACGCGGTCCAAGCAACCAACAAATCGAGGAGTCCTCGTGATTCGATTTC 1611
                |||||||||| || ||  || ||||| ||||||||||||||||||||| |||||||||||
Db    52860 CGCTAAACTCGGCGCGGTCCAAAAGCACCAACAAATCCGGAGTCCTCGTGATCCGCTACTTTC 52919

Qy     1612 AGGCCTCGGTTGCCAATGTTTTTACTCAGGCACGTGTCTAACACAG 1671
                |||| |||||||||||||  ||| |||  ||| |||||||||||  
Db    52920 ACCCCTCCGGTTGCCAACGTTTTTACTCTAGGAGTCACATGCACCGGTCTAACACCG 52979

Qy     1672 ACATGAAGATGATAGATTCAGCGGCTTATAGGATTCTCCGCCTTACAAAAGGAAAGAATG 1731
                ||||||||||||||||||||||  ||||  ||||||||||||||||||||||||||||||
Db    52980 ACATGGAAGATGATAGATTCCGCGCTTAGGAATTCTCGTGTACAAAGCGGAAAGAATG 53039

Qy     1732 TCCTGAGAGGAGATCTTAACGACTCACCAGTAATGCCTCGTTGTTGATCCATCGTGTGTG 1791
                ||||||||||||||||||||||||||||||||  || |||||||||||||||||||||||
```

```
Db  53040  TCTTGAGAGGCGATCTCAACACGCACCACTAGTCATGCCTCTGTTGATCCACCTGTGTG  53099

Qy   1792  GGTACATGGAAGTGGGCGCTGACAAACGGCAAGAGAAGGTAAGAGAGCACTATGCGTGG  1851
                ||| || |||||   || ||||||||||||||||||||||||| ||| |||||||
Db  53100  GATATAGGAAGTGGGCGTTGACGAACGACAAGAATAAGGTGACAGGAGCACTATGCATGG  53159

Qy   1852  GTGGCGGTTGTGACGCCAATTCTGAAAGTTGTGTGGAGTTCCGCTCAACGGAAGTAGGGTTAG  1911
                |||||||||||||||||||||||| || |||||| ||||||||| |||||||||||
Db  53160  ATGGCGGTTGTGACGCCAATTCTGAAGTTGTGTGGAGTACCGCTTAAGGAACCAGGGTTAG  53219

Qy   1912  CACCGGAGAATGATGATGGACTTGGATCACTTGCGCCGATGTGAGTTCTCTGAGTTGACATGG  1971
                ||||||||||||||||||||||||||| || |||||||||||||| |||||||||||
Db  53220  CACCGGAGAATGATGATGGACTTGGATCACTTGCGCCGATGTGAGTTTCTTGGAGTTTGACATGG  53279

Qy   1972  TTGGCGACTTTCACCGGTACACGGTTCGAGCATTCATCGATTAGAATCGCCAACATTCTTT  2031
                |||||||||||||||||||||||||| |||||||||||||||| ||||||||||||||
Db  53280  CTAGCGACTTTCACCGGCTACACGGTTCGAGCATTCATCGATTAGAATCGCCAACATTCTTC  53339

Qy   2032  TCCCCTGCATTTACGCTACTAGGATTCTCGAGGGCAGGAACATTGACTTCAAGCCTGCGC  2091
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53340  TCCCCTACATTGCATTTGTCATTGTGAGGGCAGGAACATTGACTTCAAGCCTCGCGC  53399

Qy   2092  TTGAAGAATCTTTATTTCGAGGCAGTCGCCGCCAACTGCAGGAGATTAGTCACACCGAGGGAG  2151
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53400  TTGAAGAATCTTTATTTCGAGGCAGTCGCCGCCAACTGCAGCAGATTAGTCACACCGAGGGAG  53459

Qy   2152  CTACAATAGAGAGATGTTGATGAGACATATGATAGATAGGCGGAGTTTGACACGAGCA  2211
                |||||||| ||| ||||||||||||||||||||| ||||||| || ||||||||||||
Db  53460  CTATAACAGAGAGATGTTGATGAGACAGAGACAGATTAGGCGGACTTTGACACGAGCA  53519

Qy   2212  TGTATCATTTCAGTGAGCATATACTTCCAGCGAGGAAAAGCAAGAGTTTGAGCGGAAGCTC  2271
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53520  TGTATCATTTCAGTGAGCATATACTTCCAGCGAGGGGAAAGCAAGAGTTTGAGCGGAAGCTC  53579

Qy   2272  ACAGGAACAACAGCAAGCTGCAGAGTGGTGCAAGAAACAGCAGGATAAGTACTGCGCCAAGT  2331
                |||||||| ||||||||||||||||||||||| |||||||||||||||||||||||||
Db  53580  ACAGGAATAACATCAAGCTGCAGAGTGGTGCAAAAACAGGATAAGTTTCTGCGCCAAGT  53639

Qy   2332  GCCTCAGGGCTATCAAGTTTCTGAAGGACAGAGATCAGCTGCTCCTCTTCCACTACAACTA  2391
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53640  GCCTCAGGGCGTATCAAGTTTCTGAAGGACGATAAGCAGCTGCTCCTCCTCACTTCCACTACAGCTA  53699

Qy   2392  TTCCGCAATGACACAGCTCCCTCAGGACATGCCTTCGAGCGAGATATGACGCGCCCC--------  2444
                ||||| |||||||||||||||||||||||||||||| ||||||||| ||||||||||
Db  53700  TTCCACAGGGACAGCTCCCTCAGGACATGCCTTCGAGCGGAGGATATGACGCGCCCGAGCCAA  53759

Qy   2445  ----------GAGCCTAGAGCAGAGATTCGATGTCCCTGCCATGTCCTGCGAGGCATTCAT  2490
                            |||||||||||||||||||||||||||||||||||||||||||
Db  53760  TTCGGCGCAAACCTGAGCCTAGAGCAGAGATTCGATGTCCCTGCCATGTCCTGCGAGGCATTCAT  53819

Qy   2491  CATTCGAGCCTCGTGAATCTAGGAAGCATTAGGAGAACGACACTCAGTCGATCTAGCAGCA  2550
                ||||||||||||||||||||||||| ||| |||||||||||||||||||||||||||||
Db  53820  CATTCGAGCCTCCATGAATCTGGGAGGAAGAGGAGGAGACCACACTCACTCAATCAATCTAACAACA  53879

Qy   2551  GGAGCAGAGACACTTCTGCAGTCTCGTAGTTTACGGCGACCGCGGTTGCTCGGCCGCAATAGAA  2610
                |||||||||||||| || ||||||||||||||| |||||| ||||||||||||||
Db  53880  AGGGTAGACCGACACTTCTGCATTTCGCCATTCTCGTAGTTTAC----------------ATGAA  53917

Qy   2611  GAAGAGAGGTCGAGTATCCTCAGCAGCGGTCGCCGCCGGCCACAGAGCTGATGAGATCGAGT  2670
                |||||||||||||||||||||||||||||||||||||||||| || ||||||||| |||
Db  53918  GAAGAGAGGTCGAGTATCGTCTCAGCAGCGGTGATGCCGCCCACAGAGCTGAGGAGTCGAGT  53977

Qy   2671  ACCCACATGCTGGAGCTGATACGGAACATGCCGGTTCGTTATGGCTTGGGAGCAATCAC  2730
                |||||||| ||||||||||| |||||||||||||||| || |||| |||||||||||||
Db  53978  ACCCACCTACTCTGGAGCTGATACGGAGCAGGGCCGGTTCGTCATGGCCAGGGAGCAATCAC  54037

Qy   2731  AGGCAGCCATTGACTACCAACTTCGTTCATTATTCGACTTCGTTCATTCTTCGACTTCACTTCAC  2790
                |||||||||||||||||| |||||||| |||||||||||||||||||||||||||||
Db  54038  AAGCAGCCATTGACGACCGACTTCGTTCATTCTTCGACTTCGTTCATTCTTCACTTCAC  54097

Qy   2791  CATTATATTATATCCTTGTCTGTGATTGTCTTCTTTATTTTGTTTCTTATTTTGTTTCAGTGATTGGATTTGT  2850
                |||| ||||||||||||||||||||| ||| |||||||||||||||||||||||||||
Db  54098  CATTGTATTATATCATCCTTGTGATTTGTGATTTGTCTTCTTTATTTTGTTTCTTTATTTTGTTTCTGTGAATGGATTTGT  54157
```

```
Qy   2851  CCTGAGTACTCTCTTCCAAGTTTATTCACCACAGTGGACTGTGTGATTTAAGTTTGGGGA  2910
                ||| ||||||||||||||||||||||| ||||||||||||||||||||| |||||||||
Db  54158  CCTGGGTACTCTCTCTTCCAAGTTTATTCACCACAGTGGACTGTGTGATTTGATTTAAGTTTGGGGA  54217

Qy   2911  GGGGTCAGGAAGTATG--TTGCATTGTGTATATATATTTAAGTCTGCATTCCATCTAAGGCAT  2968
                ||||||||||||||||   ||| ||||||||||| ||||| |||| |||||||||||||
Db  54218  GGGGTCAGGAAGTGTGTTATTGCATTGTGTATATTTTTTAGTCCGCCATACATCTAAGGCAT  54277

Qy   2969  AGAAAAACCAAAAAAAATTAAAAAATTTCAGAAAAATGATTCAC-AAAAAAAGAGTGTTC  3027
                ||||||| ||||||||||||||| ||||||||||||||||| |||  ||||||||||||||||
Db  54278  AGAAAAACCAAAAAAATTTGAAAAATTTCAGAAATGATTTCACCAAAAAAAAAGAATGTTT  54337

Qy   3028  ATGTAGTTGCATTACATTTAGGATCAAGTCTAGAGTGTTCATTTAGAGTTGTTGCATAT  3087
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54338  ATGTAGTTGCATTACATTTAGGATCGAGTCTAGAGTGTTTCATTTAGGATTTAGTTCGCATAT  54397

Qy   3088  GCATAGGGGATAATGATGAGAGTAGCCTTGTAAGCA  3122
                ||||||||||||||||||||||||||||||||||||
Db  54398  GCATAGGGGATAATGATGAGAGTAGCCTTGTAAGCA  54432
```

RESULT 6

```
AAF22296/c
ID   AAF22296 standard; DNA; 129021 BP.
XX
AC   AAF22296;
XX
DT   20-MAR-2001 (first entry)
XX
DE   BAC containing repeats from centromeres 1-4 #19.
XX
KW   Centromere; microsome; vector; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200055325-A2.
XX
PD   21-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-US007392.
XX
PR   18-MAR-1999;          99US-0125219P.
PR   01-APR-1999;          99US-0127409P.
PR   18-MAY-1999;          99US-0134770P.
PR   13-SEP-1999;          99US-0153584P.
PR   17-SEP-1999;          99US-0154603P.
PR   16-DEC-1999;          99US-0172493P.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Preuss D, Copenhaver G, Keith K;
XX
DR   WPI; 2000-587529/55.
XX
PT   Recombinant DNA construct comprising a plant centromere, useful for
PT   producing stably inherited microsomes which can serve as vectors for the
PT   construction of transgenic plant and animal cells.
XX
PS   Claim 102; Page 686-716; 1449pp; English.
XX
CC   The present invention relates to a recombinant DNA construct of a plant
CC   (Arabidopsis thaliana) centromere. The constructs are useful for
CC   producing stably inherited microsomes which can serve as vectors for the
CC   construction of transgenic plant and animal cells expressing selected
CC   proteins such as hormones, enzymes, interleukins, clotting factors,
CC   cytokines, antibodies, and growth factors
XX
SQ   Sequence 129021 BP; 42091 A; 22610 C; 24583 G; 39735 T; 0 U; 2 Other;

Query Match             45.3%;  Score 1415.6;  DB 3;  Length 129021;
Best Local Similarity   86.3%;  Pred. No. 0;
Matches 1610;  Mismatches 204;  Conservative 0;  Indels 52;  Gaps 2;
```

```
Qy    744 AAAGTAAAAAAAAAAAAGAAAAAAAAGTCAAAACCGACGTTTCACTCAACCGGTTTACT 803
          |||| ||||||| |  |  |||||||||||||||   ||||||||||||| |||||||||
Db  39545 AGACTCGACACAACTCACAATCCAAAAAAAAAATATTCAAAACCGACCGGTTTACT 39486

Qy    804 CGACCGGTATACACTGTTGGATTGTGTTTTTGTTATTACTAACATATTAACGTT 863
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  39485 CGAGCGGTATACACTGTTGGATTGTGTTTTTGTTTTTGCTGGAACTAACGTATTAACGTT 39426

Qy    864 TATCTTTGAGTTTGTCTGTTTTAGGTTTCATCATGAGTAACTACAGTGGAAAATCCTC 923
          |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db  39425 TATCTTTGAGTTTCATCGTTTCATCATGAGTAACTAGTGGAGAATCCTC 39366

Qy    924 TATGGACCCTGATATAATGTGAATGAAGCTAAGTCCTGGTCCACTAGACCGGAGTGAGA 983
          |||||||||||   |||| |||||||||||||||||||||||||||||||||||||
Db  39365 TATGGATCCTAATTATAATATGAGTGATGAAGGCTAGTCTTGGTCCCACTAGACCCGAGAGAGA 39306

Qy    984 GCAACATGTTTACGGAGAGCTATAGGCGATGAATTTGAACGCTCTGCAGCTTGACGTAATCA 1043
          |||||||||||||||||||||||||||||||||||||||| ||||| |  ||||||||
Db  39305 AGAACACGTTTACGGAGAGCTTTAGAGGAGAAATGAACGCTCCACAGCTGTACAACGTAGTCA 39246

Qy   1044 AAGAAGAGCTGAAATCGCTAGAGGAAAGAGGGCGATGTCGAGTAGAGTATAGAGCTCGATTGA 1103
          ||||||||||||||||||||||||||||||| |||||||||||||||||| |||||||| |
Db  39245 AGGAAGAGCTGAAATCGCTAGAGGAAGAGAGAGGCAGTTTCGAGTAGAGTATGATGAGTGTTAATTGA 39186

Qy   1104 TGAGGGATATCAAAACTGAGTATGAGCCAGAGTCATGGCGCCAAGGAGACCGAAGCTACTGAA 1163
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  39185 TGAGGAGATGTGATAAGACAAGAGTGTGATAAGGCGCGCAAGGAGACAGTGCTGCTGAA 39126

Qy   1164 CAAATCCGACGCAGTTACAGTGGAGAGTATATCAGATTCTTTGAGATGAATGACTTCTG 1223
          |||||||||||||||||| ||||||||||||||||||||||||||||| |||||||||
Db  39125 CAAACCGACGCAGTTACAGTGGAGAGTATATCAGATTCTTCGAACATAATGATTTCTG 39066

Qy   1224 GGGAACGAGGTATCCCGATATGAGACTTTAGCCCAGTTGGGGTTACTGGAGGACGTGCA 1283
          ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db  39065 GGGAACGAGGTATCCCGTTTATGAGACTTTAGCCCAGTTGGGGTTACTGGAGGATGTGCA 39006

Qy   1284 GCATCTGTTCGAGAAGTGTCATCTG---ATAA 1312
          ||||||||||||||||||||||||
Db  39005 ACATGTGTTCGAGAAGTGTCATCTGAGAACACTGTCTTGAGACACACTGTCTTATAA 38946

Qy   1313 GGAGGAGACAATCGAGTTTCTTCTTTCCACACTGCAAGTGGAAATGTATGAGGAGAGAT 1372
          ||||||| |||||||||||||||||||||||| ||||||||||| |||| ||||||||
Db  38945 GGAGGAGACAATCGAGTTTCTTCTTTCCACACTGCAAGTAGAGATGTATCAGGGGTCTCACAGA 38886

Qy   1373 CTTTGAGCTGGATACCATGGGGTTAGGCTTCTTGAGCGTTCTTAGTGGATGAACAGCGGTA 1432
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db  38885 CTTTAAGCTGGATACCATGGGGTTAGGGTTCTTCGAGCGTTCTTCAGTGGATGAACAGCGTTA 38826

Qy   1433 CCAGATTTAGATCAAGAAATTGGAAGGAGTGGGGAAAGGGAACCAA 1492
          |||||||||||||||||||||||||||||||||||||||||||||
Db  38825 TGAGCTGTCGATCAAGAAGAGCTTAAGGAGTGGGGAAAGGGAACCAA 38766

Qy   1493 CCCCAGGTTGACAGCCGTCCAAGAGCAACCAAATCCGGAGTCCTACTTTCA 1552
          |||||||||||||||||||||||||||||||||||||||||||||
Db  38765 CCCCAGGTTGACAGGGAAGAGCTTAAGGAGTTGTGGGCTACTATTGGGAACAATGTACC 38706

Qy   1553 GCTAAACTCGACGCGGTTCCAAGAGCAACCAAATCCGGAGTCCTGTGATTCGCTACTTTCA 1612
          |||||||||||||||||||||||||||||||||||||||||||||
Db  38705 GCTAAACTCTGCGCGGTTCCAAGAGCAACCAAATCTAGAGTCCTGTGATTCGCTACTTTCA 38646

Qy   1613 GCGCTCGGTTGCCAATGTTTTTTACTTCCAGGGAGTCTACAGGCACCGGTCTTAACACAGA 1672
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  38645 GGGCTCGGTTGCCAATGTTTTTTATTCTGCAGGGAGTCTGCAGGCACCGGTCGTCTTAACGGA 38586

Qy   1673 CATGAAGGATGATGAGATTCAGCGTAGCTTAGCGGCTTACACGGATTCTCGGCCTTACAAAGGAAAGAATGT 1732
          |||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db  38585 CATGGAATGATGATGATTTCCGCACTTTACAGGGATTCTCCGGCCGTACAAAAGGGAAAGAATGT 38526

Qy   1733 CCTGAGGAGGAGATCTTAACGACTCACCACCACCAGTAATGCCTCTGTTGATCATCTGTGTGG 1792
          ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db  38525 CTTGAGGAGGAGGATCGGGATTCTCAACGCAATGGGGATTCTCATTGATCACCTGTGTGG 38466

Qy   1793 GTACATGAAGTGGCGCTGACAAACGGCAAGAAGAAGGTAAGAGAGAGCACTATGCGTGGG 1852
          ||| ||||||||| ||| |||| ||||||||||||||||||||||||||| ||||||
Db  38465 ATACAGGAAGTGGGCGGTTGATAAACGGCAAGAAGAAGAAGTAACAGGAGCACTATGCGTAGG 38406

Qy   1853 TGGCGGTTGTGACGGCCAATTCGAAATTGTGGAGTTCCGCTCAAGGAAGTTAGC 1912
          |||||||||||||||||||||||||||||||||||||||||||||| |||||
Db  38405 TGGCGGTTGTGACGCCAATTGTGGAGGCTGTTGGAGTACCTTCAGGAACCAGGGCTAGC 38346

Qy   1913 ACCGGAATGATGACTTGGATCACTGCGCGCGATGTGAGTTCTGAGTTGACATGGT 1972
          |||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db  38345 ACCAGAATGATGACTTGGATCATTGGGCCGATGTGAGTTCTTGGAGCATGACATGGT 38286

Qy   1973 TGGCGACTTTCACCGTTCAGCAGGTTCGAGCATTCATCGATTAGAATCGGCCAACATTCTTTT 2032
          ||| ||||||||| ||||||||||||||||||| |||||||||||||||||||||
Db  38285 TGGTGACTTTCCATCGCTACAGGTTCGAGCATTCATCGATCAGAATCGGCCAACATTCTTCT 38226

Qy   2033 CCCCTGCATTTACGCTACTAGGATTCTGCAGCTTGACGTAATCA 2092
          ||||| ||||||||||||||||||||||||||||||||||||
Db  38225 CCCCTACATTGACGCTACTAGGGTTCTCGCAGCTGACGTAATCA 38166

Qy   2093 TGAAGATCTTTATTTCGAGGGCAGTCGCATGTCGAGTAGATATGAGCTCGATTGA 2152
          |||||||||||||||||||||| |||||||||||||||| ||||||||||
Db  38165 TGAAGATCTTTATTTCCAGGGCAGTCGCATGTCCAGTCAGTCAGATCAGTCAGTCACACCGAGAGAGC 38106

Qy   2153 TACAATAGAACAGTGTGATGAGACATATGATATGATAGAGATGAGCCGGAGTTTGACACGCAT 2212
          ||  | |||||||||||||||||||| |||||| ||||| ||||||||||||||||||||||
Db  38105 TACGACAGAAGAGTGTTGATAGACAGATGAAATAGATGAGGGGGAGTTTGACACGAGCAT 38046

Qy   2213 GTATCATTCAGTGAGCATATACCTCCAGCGGAAAAGCAAGAGTTGAGCGGAAGCTCA 2272
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  38045 GTATCATTTCAGTGAGCATGTGACATGCCTCCAGCAAGGGAGCAAGAGTTGAGCGGAAGCTCA 37986

Qy   2273 CAGGACAACAGCAAGCTGCAGAAGTCGTGCAAGAGTGGTCAAGAACAGGATAAGTTACTCCGCCAAGTG 2332
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db  37985 TAGGAAAAAACAGCAAGCTGCAGAAGTGGTCAAGAACATGATAAGTTACTCGCCAAGTG 37926

Qy   2333 CCTCAGGGCTATCAAGTTTCTGAAGGACAAGATCAGTCGTGCTCCTCTTCCACTACAACTAT 2392
          ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db  37925 CCTCAGAGCTATCAAGTTTCTGAAGGAGGATAAGCTAAGCTGTCTCCTCTTCCACTACAGCTAT 37866

Qy   2393 TCCGCAATGACAGCTCCCCTCCAGGACATGCGCCTTCGAGGAGATATGACGC-------- 2440
          |||||||||||||||| ||||||||||||||||||| || |||||||||||||
Db  37865 TCCGCAGGGCGACGATCCCTGCGACATGCCCTGACATGCGCCTGAGCGGAGCGGAGAAG 37806

Qy   2441 --------GCCCCGAGCCGAGCTGAGCGGAGCCGAGCGCGGCGATTCATC 2491
                  |||||||||||||||||||||||||||||||||||||||||||
Db  37805 TCGGCGCAGGCCTCGAGCCTGAGCGGAGGCTGAACAGGAGAGTTCCGAGGAGAGTTCCGATTCATC 37746

Qy   2492 AATTCGAGCCTCGTGAATCTAGAAGGAACGACACTCACTCGATCGATCTAGCAGCAG 2551
          ||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db  37745 AATTCGAGCCTCGTGAATCTGGGAGAAACGAGGAGAGAGGAGACCCACTCACTCGATCTAGCAGCAA 37686

Qy   2552 GAGCAG 2557
          ||||||
Db  37685 GAGCAG 37680
```

RESULT 7
AAF30458

PD   22-MAR-2001.
XX
PF   14-SEP-2000; 2000WO-EP008994.
XX
PR   16-SEP-1999; 99GB-00021964.
XX
PA   (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI   Steimer A, Mittelsten Scheid O, Paszkowski J;
XX
DR   WPI; 2001-244815/25.
XX
PT   Novel transcriptionally silenced genes obtained from Arabidopsis useful
PT   for selecting plants which compared to wild type plant are impaired in
PT   transcriptional gene silencing.
XX
PS   Claim 1; Page 26-28; 36pp; English.
XX
CC   This present sequence is that of a 'virtual' cDNA combining overlapping
CC   sequences from 5' and 3' extensions (see AAF30450-53) of an Arabidopsis
CC   thaliana transcriptionally silent information (TSI) sequence, i.e. TSI-A
CC   (see AAF30454).The 4860 bp sequence was used in Arabidopsis genomic
CC   database searches, identifying a genomic sequence (see AAF30459) that
CC   included a transcribed region. Comparison of transcriptional gene
CC   expression between an Arabidopsis line carrying a silent transgene
CC   present in multiple copies and its mutant derivative mom1 impaired in
CC   silencing of the transgene revealed 2 cDNA clones which were expressed in
CC   the mutant plants, but not in the parental or wild-type plants. Both
CC   clones were derived from the same family of transcripts referred to as
CC   TSI. Genomic templates encoding TSI are repetitive elements with mainly
CC   pericentromeric location and conserved organisation among various
CC   ecotypes. Transcriptional silencing of genomic TSI templates is
CC   specifically released in the mutant. Transcription of TSI can be used as
CC   a marker to identify a defective silencing pathway in a plant. The
CC   invention provides a nucleic acid encoding genetic information which is
CC   silenced in wild-type plants but whose expression is turned on in plants
CC   defective in transcriptional gene silencing, and has the formula: RA-RB-
CC   RC, where RA, RB, RC constitute component nucleotide sequences selected
CC   from the present sequence and those given in AAF30450-59. Also provided
CC   are methods of selecting a plant that is impaired in transcriptional gene
CC   silencing, and of producing DNA representing at least part of a gene
CC   necessary to maintain silencing of another gene in a cell or plant
XX
SQ   Sequence 4860 BP; 1296 A; 1068 C; 1131 G; 1365 T; 0 U; 0 Other;

Query Match   44.8%; Score 1398.6; DB 4; Length 4860;
Best Local Similarity   76.2%; Pred. No. 0;
Matches 1836;   Conservative 0; Mismatches 489; Indels 84; Gaps 6;

Qy   788 ACTCAACCGGTTACTCGACCGGTACGCTGGTTGAGTTGTCGTTGTGTTTTGGTTTGCTATTA 847
             ||||| ||| ||||||||| ||||| |||| ||||||| ||| |||||||| ||||||||||
Db   328 ACTCGACCGTTCACTCGATCGCGTACTTGACT--GGTTTAGTGTGTGTTTATTTGGAA 385

Qy   848 CTTAACATATTAACGTTTATCTTGAGTTC--GTCTGTTTTTAGGTTTACATCATGAGTAAC 906
             |||| ||||||| |||||| ||||||||| |||||||||||||||| ||||||||||||
Db   386 CTTAACATATTGATATTTGAGTTTGAGTTTACATTCTTTTTCAGGGAATCAATATGAGCAAC 445

Qy   907 TACAGTGAAAATCCTCTATGGACCCTGATTATAATGTGGATAAGAGCTAAGTCCTGGTCC 966
             ||||| |||||||||||| |||| ||| ||| |||||||||| ||||||||||||||
Db   446 TACAGTGGCGAATCCTCCATGGATCGGATTACAACGTCGAATCTTGGTCA 505

Qy   967 ACTAGACCGGAGTGAGAGCAACATGTTTACGAGAGCTTACGAGAGGGATGAATTTGAACGCTCT 1026
             ||||||||| ||||||||||||||| |||||| ||| |||||||||||||||||
Db   506 ACTAGAACCAGAGAGAGAGCAACAGCTTATGAGAGCTTCAGAGCGCCGAGACCCAACGCTCA 565

Qy   1027 GCAGTCGACGTAATCAAAGAAGAGCTGAAATCCGCTAGAGAAGAGACGGCGATGTCGAGT 1086
             || |||||| |||||||||||||| ||||||||||||||| ||||||
Db   566 GTAGCTCGACGCAATGAAAGGAGAGCTGAGATTGCTAGAGAAGAGAGCGATGACCAGC 625

Qy   1087 AGATATGAGCTGATGATGGAGGATATCAAAACTGAGTATGAGCCAGAGTCATGGCGCAAG 1146
             ||||||||||| ||||| ||||||||||||||||||||||||||||||||||
Db   626 AGATATGAGTTGATCGACGAAGATATTGACGTCAGAGAGTATGAGCCTGAGTCATGGCGCACAGA 685

Qy   1147 GAGACGAAGCTACTGAACAAATCCGACGAGGTTACAGTGAGGAGTATATCAGATTCTTT 1206
             ||| || ||| ||||||||||||| |||||||||||||||||||| ||||||||||||
Db   686 GAAACAAAACTGTTGAACAAGCCTGATGAAGTTACAGTGAAGAGTACATCAGACTTTTC 745

Qy   1207 GAGATGAATGACTTCTGGGAACGAGGATCCTGATATGACTTAGCCCAGTTGGGG 1266
             |||| ||||| |||| |||||||| |||||||||||| | |||||||||||||||
Db   746 GAGCTGAACGACTTCTGGGGAGCGAGGTACCCCGTTTATGAGACTCTAGCCCAGCTTAGG 805

Qy   1267 TTACTGGAGGACGTGCAGCATCGTCTTCGAGAAGTGTCAT--------- 1305
             ||||||||||| ||||||||||||| |
Db   806 CTACTGGAGGACGTACAGCACTTATTCGAGAAGTGCCATCTTCTTGAGACGCTGATGTCTAC 865

Qy   1306 ----------CTGATAAGGAGGAGACAATCAGTTTCTTCCACACTGCAAGTGGAAATG 1355
                       ||||| ||| ||||| ||||| ||| ||| |||
Db   866 CCGTACGTCGCTTACAGAAGGAGAAACAATAGAGTTTCTCCTCCACTCGCAAGTGGAGTTG 925

Qy   1356 TATGAGGGACTCACAGACTTTGAGCTGATACCATGGGGTTAGGCTTCTGACGTTCTA 1415
             ||||||||||| ||||| ||||||||||||| |||| |||||||| |||||||| ||
Db   926 TATCAGGGACTTACTGCAGATGAACTGGAGAGTGAAGGTTGGGATTCTTGACTTTTTCA 985

Qy   1416 GTGGGATGAACACGGGTACCACCCAGATTTAGATCAAGAAAATTGGAAGAACTGTTTGGTTCCCT 1475
             || ||||||||| || |||| |||| ||| |||||| |||| |||||||||||||| |||||
Db   986 GTTAACAGCAGCAGCGTTACCAGCTATCATCAGAGCGTTGGAAGGATTATTTGGTTTTCCC 1045

Qy   1476 AGTGGAAAGGGAACCAACCCCAGGTTTGACAGGGAAGAGCTTAAGGAGATTGTGGGCTACT 1535
             ||||||||||||||||| |||||||||||||||||||||| |||| ||||||||||| |
Db   1046 AGTGGAAAGGGAACTAAACCCAAGTTCGAAAGGGAAGAGGAAGGATTGTGGGTTAACC 1105

Qy   1536 ATTGGGAACAATCTACCGCTAAACTCGACGCGTCCGTCGCCAATCCGGAGTCCT 1595
             || |||||||| |||| |||| |||||||| |||||||| ||||||||||||||
Db   1106 ATTGGGAACGATTTGGCCTCAACTCGCAAGTCTAAGAGCATCTAAGACCAGATTCGAAGCCCT 1165

Qy   1596 GTGATTCGCTACTTCAGCGCTCGCGGTTGCCAATGTTTTTACTCCAGGGAGTCTACAGC 1655
             |||||||| |||| ||||||||| || |||||| ||||||||||| ||||||||||||||
Db   1166 GTGATCCGCTACTATCAGCGCTGCAGTGGCGGTGGTTCTGTACCCCAGGGAATCTACAGC 1225

Qy   1656 ACCGTGTCTAACACAGACAGACATGAAGATGAGATCAGGGCTTATAGGGATTCTCCGCCTT 1715
             |||||| |||||||| |||||||||||||||||| ||||||||||||||||| |||| |||
Db   1226 ACCGTGTCTAACACAGACAGACATGAGGATGATTAGATTATCTGCACTCAAGGGTATTCTCCGGAGA 1285

Qy   1716 ACAAAAGGAAGAAATGTCCTGAGGAGATCTTAACGACTCACCACCAGTAATGCCTCTG 1775
             || ||||||||| |||| ||||||| ||| |||||| |||||||||| |||||||||||
Db   1286 ACAAAGGGAAGAAGGTTCCTAAAGGGCACCTTAATGATACACCACCGGTCATGTCTCTG 1345

Qy   1776 TTGATCCATCTGTGTGGGTACATGAAGTGGGCGCTGACAAACGGCAGAGAAGGTAAGA 1835
             ||||| ||||||||||||| ||||||||||||| |||| |||||||| ||||| |||||
Db   1346 TTGATCCATATGTGTGATCAGGAGTGGCGGTGTGAGTGACACCAACCGGGAGGAAGGTGCGA 1405

Qy   1836 GGAGCACTATGCGGGTGGGGTGGCGTTGTGACGCCAATTCGAAGTTGTGGAGTTCCGCTC 1895
             |||||||||| ||||||| || ||||| || || ||| || |||||||||||| ||||||
Db   1406 GGAGCCCTTGTCTGGTGGGGTGTGGGGTGTGACCGTGATCTCGATTCGATGCGATGGTGTACCCTC 1465

Qy   1896 AAGGAAGTAGGGTTAGCACCGAGAATGATGAGCTTGGATCACTTGGGCGGCGATGTGAGTTC 1955
             || |||| |||||||| ||||||| ||||| |||||||||| |||||| |||| ||||
Db   1466 ACGTCTCCAGGGTTGATCCGAGGATTAGCACTTTAGATTAGATCGATCTTGGTGTGAGTTT 1525

Qy   1956 TCTGAGTTTGACATGGTTGGCGACTTTCACCGTCCAGGTTCGATGCATTCATCGATTAGA 2015
             |||| |||||||||| |||||| |||||||| ||||||||||||||||| |||||
Db   1526 CTGGAGTACGACATGGTTGGCGATTTCTCCTATCGCTACAAATCGAATCTTGGTCA 1585

Qy   2016 ATCGCCAACATTCTTTCCCTGCATTTACCGTCATTTACGGTACTTGAGATTCTCGAGGGACATT 2075
             || |||| |||||| |||| ||||| ||||| ||||||| || ||| ||||||||| ||||||
Db   1586 ACAGCCAACATTTGCTTCCCTGCACCATTTAGGGTGAGAACATT 1645

Qy   2076 GACTTCAAGCTGCGCTTGAAGATCTTATTTGTGATCGTTTCATCATGAGTAAC 2135
             |||||| ||||| ||||||||||||| |||| |||||||| |||||||||||||
Db   1646 GACTTCAGACCTGCGCGGTGATTACCTCTACTTTGAGAGCACTCGACTGATGACAAT 1705

Qy   2136 AGTCACACCGAAGAGCTACAAATAGAAACTGAGTATGAGCCAGAGTCATGGCGCGCAAG 2195
             |||||||||| ||||||| ||||||||| |||||||||||||||||||||||||||||||
Db   1706 GTCCCTACGACCGGAAGCTACAGAGGATGATGAGTTTGCTGAGCGGATGAGATTTTGCTGAGCAAG 1765

Qy   2196 GAGTTGACACGAGCATGTATCATTTCAGTGAGCATATACCTCCAGTCGGAGAAAGCAAG 2255
Db   1765

```
XX
DT   20-MAR-2001  (first entry)
XX
DE   BAC containing repeats from centromeres 1-4 #3.
XX
KW   Centromere; michrosome; vector; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200055325-A2.
XX
PD   21-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-US007392.
XX
PR   18-MAR-1999; 99US-0125219P.
PR   01-APR-1999; 99US-0127409P.
PR   18-MAY-1999; 99US-0134770P.
PR   13-SEP-1999; 99US-0153584P.
PR   17-SEP-1999; 99US-0154603P.
PR   16-DEC-1999; 99US-0172493P.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Preuss D, Copenhaver G, Keith K;
XX
DR   WPI; 2000-587529/55.
XX
PT   Recombinant DNA construct comprising a plant centromere, useful for
PT   producing stably inherited michrosomes which can serve as vectors for the
PT   construction of transgenic plant and animal cells.
XX
PS   Claim 102; Page 336-351; 1449pp; English.
XX
CC   The present invention relates to a recombinant DNA construct of a plant
CC   (Arabidopsis thaliana) centromere. The constructs are useful for
CC   producing stably inherited michrosomes which can serve as vectors for the
CC   construction of transgenic plant and animal cells expressing selected
CC   proteins such as hormones, enzymes, interleukins, clotting factors,
CC   cytokines, antibodies, and growth factors
XX
SQ   Sequence 67087 BP; 20469 A; 12665 C; 12858 G; 20613 T; 0 U; 482 Other;

Query Match             43.8%;  Score 1366.8;  DB 3;  Length 67087;
Best Local Similarity   75.9%;  Pred. No. 0;
Matches  1832;  Conservative    0;  Mismatches  492;  Indels  90;  Gaps  8;
```

Sequence alignment (Qy = Query, Db = Database):

```
Db   1766  GAGTATGATACGAGCATGTATCATTTCAGTGAGCACGTACCTTCCAGCGCAGGAGAGCAAG  1825
                  ||||| ||||||||||||||||||||
Qy   2256  AGTTGAGCGAAGCTCACAGGAACAACAGCAAGCTGCAGAAGTGGTGCAAGAAACAGGAT   2315
            |||| |||| ||||||| || |||||| ||||| ||||||  |||||||||||| ||||||
Db   1826  AGCTTGAGTGAAGCTCACAGAGAAACAACAGTAAGTTGCAGAGAGGTGGTGCAAGAAACAAGAT 1885

Qy   2316  AAGTTACTCCGCCAAGTGCCTCAGGGCTATCAAGTTTCTGAAGGACAAGATCAGCTGCTCC  2375
            ||||||| | ||||||||| ||||| | ||| ||||||  ||||||||||||  || ||||
Db   1886  AGGTTACTTATCAAGTGCTTCAGGCCATCACGTTTCTAACGGACAAGATAAGTTGCTTC    1945

Qy   2376  TCTTCCACTACAACTATTCCGCAGACGTCCCTCAGGACATGCCTTCGAGGAGATAT   2435
            ||||| || || ||||||||||| |  | ||||||||||||||||||| |||||||
Db   1946  TCTTCTACCACAGCTATTCCGCAGGGAGCGTCCTCAGGACATGCCTTCGAAGAGATAT   2005

Qy   2436  GACGC------------------------------GCCGAGCCTAGA   2453
            |||||                            || |||||||||||
Db   2006  GACGCGCCCAGGGGCCAAGTCATCACAGCGCCTGAGCCTGAGCCTAGT   2065

Qy   2454  GAGCAGAAGATTCTTGCATGTCCCTGCGAGGCATTCATCATTCGAGCCTCGTGATCTAGG  2513
            ||| | |||||| | || ||||||| ||||||||||||||  |||||| |||| | ||||
Db   2066  GACCGAGTAGTACCACCAGTCCCTGCAAGGCATTCATCATTCGAGCCTCGGG    2125

Qy   2514  AAGAATAGGAGAACGACACTCACTCGATCTAGCAGCAGGAGCACGAGCACTTCTCCGAGTCT  2573
            || | |||||||| |||| || |||| || ||||| |||||||||||||||||| || |||
Db   2126  AGAAAGAAGAAGGCTGCACTCGCTAGGTCTGCAGGAGTAGGACGACTTCTCCAGTCC   2185

Qy   2574  CGTAGTTTACGCGACCGCGTGCTGCCCGCAATAGAGAAGAGAGGTGAGTATCCTCCGAG  2633
            |||||||||||||||| |||| || |||| ||||||||||| | |||||||||||||||
Db   2186  CGTAGCTTACGCGACCGCGTGCTGCCGCCCAGCAGCAGAGAGAGAGGTGAGTATCATCAG  2245

Qy   2634  AGCGGTGCTGGCCCGCCACAGAGCTGATGAGCTGAGTACCCACATGCTGAGCGTGATACG  2693
            ||||||||||| | ||||||||||| || || |||||| || ||||||||||||||||||
Db   2246  AGCGGTGCTGCGCCGCCGGAAGGAGCAGAGTCGAGTAGCCCCAGGGGGAAGCTGAGACA  2305

Qy   2694  GAACATGGGCGGTTCGTTCGTCTATGGCTTGGGAGCAATCACAGGCAGCCATTGACTACCAACTT  2753
            ||||||||| |  |||||| ||||| |||| || |||||||| |||||||||||| || |||||
Db   2306  CAACAGGGGGAGATTCTTCGACGTGGCTTGGAGCAGCAATCACAGGCGCTATTGACGACCAACTC 2365

Qy   2754  CGTTCATTATTCGACTGAGGTAAGCGCCTCACTTCACCATTATATATTATCATCTCTTGT  2813
            |||| || || ||||||||| |||| || ||||| ||| ||||| |||||||||| |||
Db   2366  CGCTCCTTCTTCCACTGAGGTATGCACCTCACTCACTCCACCATTGTAATATACCATCTC----  2421

Qy   2814  GATTTGTTCTTTATTTTGTTTCAGTGATTCAGTGATTGTCCTGAGTACTCTCTTTCCAAGTTT  2873
               ||||| ||||| | |||||||  |||| ||||| | ||||| ||||| ||||| ||||| ||
Db   2422  ---TTGTTTTATTTTGTTTTTGTGATGTGTTTTGTTTCCTGAGTACTCTCTTTCCAAATTT    2478

Qy   2874  ATTCACACAGTGGACTGTGTGATTTAAGTTGTGGGGGAGGGCTCAGGAAGTATGTGTTGCATT  2933
            || |||| |||||||| |||| ||| |||||||||||| || |||| ||||||| |||| |||
Db   2479  GGTCACACAGTGGACTGTGTGATTTAAGTTTGGGGGAGGGCTCAGGAAGTGTGTGTTGCA    2538

Qy   2934  GTATATATTTTTAAGTCTGCATTCATCTAAGGCATAGAAAACCAAAAAAAATTAAAAAA  2993
            || || ||| ||||||||||||||||||| || ||||||||||| |||| |||||||||
Db   2539  TTGTATAATCTTGAGTTTGCATTCATCTAAGGCATAGAAAAAACC-AAAAAAATTGAAAAA   2597

Qy   2994  TTTCAGAAAATGATTTCACAAAAAAGAGAGTGTTCATTGAGTTGCATTACATTTAGGATCA  3053
            ||| ||||||||||| |||||||||||||||||||||||||||||||||| |||||||| |
Db   2598  TTTCCAGAAAATGATTTCACAAAAAAATAGAGTGTTCATTGAGTTGCATTTAGGATCG   2657

Qy   3054  AGTCTAGAGTGTTCATTTAGGATTCGCATAATGTTGCATAAGGGATAAATGATGAGATAGCC  3113
            ||| ||||||||||||| |||||||||||||||||||||||| ||||||||||||||||||
Db   2658  AGTTTTAGAGTGTTTCGTTTAGGATTTGTTCATTGTTGCATAAGGGATAATAATGAGATAGCC   2717

Qy   3114  TTGTAAGCA  3122
            |||||||||
Db   2718  TTGTAAGCA  2726
```

```
Db  18800  ACTCGACCGTAATACTCGACCGTGTACTTG--ACCGGTTAGTGTTTGCATTTATTTGGA 18857
Qy    788  ACTCAACCGGTTACTCGACCGTCACGCTCGGTTGGTTTGATTGTGTTTTTGCTTTTTGCTATTA  847

Qy    848  CTAACATATTAACGTTTATCTTTGAGTTTCGTCTGTTTTTAGGTT-TCATCATGAGTAAC  906
Db  18858  CTAACATATTGACGTTTGGCTTTTGAGTTTACATTCTTTTTCAGGAAAATCAATGAGTAAC 18917

Qy    907  TACAGTGGAAAATCCTCATGGACCCCTGATTATATGTGGATGAAGCTAAGTCCTGGTCC   966
Db  18918  TACAGTGGCGAATCCTCCATGAATGCGGATTACAACATCGATGAAGCTGGTCA   18977

Qy    967  ACTAGACCGGAGTGAGACGAACATGTTTACGAGGGATATTTGAACCGCTCT   1026
Db  19038  ACTGGACCAGAGAGAGAGCGCACAAGCTTATGAGAGCTCAGAGCGGAGACCCAACGCTCA 19037

Qy   1027  GCAGCTCCGACGTAATCAAAGAAGAGAGCTGAAATCCTAGAGAAGAAGAGGGCGATCGTCGAGT 1086
Db  19038  GTAACTTGACGCAATGAGAGAGAGGATATGAGCCAGAGTATCAAAACTGAGTATGAGCGAGTAAC 19097

Qy   1087  AGATATGAGCTGATTGATGAGATCAAAATCCGACCAGAGTTACAGTCGAGT   1146
Db  19098  AGATATGAGTTGATCGACCAGATATTGACGTCGAGTATGAGCCTGAGTCGAGT 19157

Qy   1147  GAGACGAAGCTACTGAACAAATCCGACAGAGTTACAGTCGAGTGAGTATATCAGATTCTTT  1206
```

RESULT 8
AAF22280
ID   AAF22280 standard; DNA; 67087 BP.
XX
AC   AAF22280;

```
Qy   2256  AGTTTGAGCGAAGCTCACAGGAACAACAGCAAGTGTCAGAAGTGGTGCAAGAAACAGGAT  2315
           ||||| |||||||||||||||  | ||||||||| | ||||| ||||||||||||||||
Db  20297  AGCTTGAGCGAAGCTCACAGAAACACAGCAAGTTGCAGAGGTGGTGCAAGAAACAAGAT  20356

Qy   2316  AAGTTACTCGCCAAGTGCCTCAGGGCTATCAAGTTTCTGAAGGACAAGATCAGCTGCTCC  2375
           |||||||||| |||||||| ||| ||||||||||||||||| | ||||| |||| ||||
Db  20357  AGGCTACTCATCAAGTGCTTCAAGGCCATCAAGTTTCTGACAGACAACAAGCTAAGCTGCTCC  20416

Qy   2376  TCTTCCACTACAACTATTCCGCAATGACAGCTCCCTTCGAGAGATATG  2435
           |||||  |||||   ||||||||  |  | || ||| | |||||||| ||
Db  20417  TCTTCTACCACAGGCGATTCCGCAGGAGGAGCCTCTCCAGGACAATGCCCCCGAGGAGATATG  20476

Qy   2436  GACGC------------------------GCCCGAGCCTAGA  2453
           |||||                        ||| ||||||||
Db  20477  GACGGCGCCAGAGCCAACTCGCCACAGGCCTGAGCCAAGTCACCCACATGCGCTGAGCCTAGT  20536

Qy   2454  GAGCAGAAGAGATTCTGCATGTCCCTGCGAGGCATTCATCATTCAGACCTCGTGAATCTAGG  2513
           ||||||||||||| |||||||||||| || || || |||||||||  |||||||| |||
Db  20537  GACCGAGTAGTCTTCCACAAGTCCCTGCGCATTCATCATTCAGCCTCGGGGAGCTCGGG  20596

Qy   2514  AAGAATAGGAGAACGACACTCACTCGATCTAGCAGCAGAGACAGTTCTGCAGTCT  2573
           | ||||| ||||| | | |||| | || |||  | ||||||||||||||||| ||
Db  20597  AGAAAGAAGAAGCTGCACCTCGGTCGGCAGCAGAGTACACGACTTCTCCAGTCC  20656

Qy   2574  CGTAGTTTACGGCACCGCCGGTGCTGCCGCCAATAGAGAAGAGGAGGTCGAGTATCCTCAG  2633
           || | |||| || | | || |||| ||||  | |||||||| ||||| |||| ||| ||
Db  20657  CGTAGCTTACGCGTACGCGGTGTGCTGGCCGCAGCAGAAAAAGAGAGATCAGTATCATCAG  20716

Qy   2634  AGCGGTGCTGGCCGCCACAGAGCGTCGTATGGCTTAGGCGCCTTATCACTTCACCATTATT  2693
           |||| ||||| | || ||||| ||  | ||| ||||||| ||||||| | ||||||
Db  20717  AGCGGTGCTGGCCGGAGCTAAGGACGTGAGTTTGAGTGAGCAGCCATTGACCATGTGATACG  20776

Qy   2694  GAACATGGCGGTTCGTCTATGGCTTGGGAGCATGAGTAAGCCCTGATTGTCCTGAGCAGC  2753
           ||||  ||||| ||  || |  ||| |||||||| | |||| | |||||||||||||||
Db  20777  CAACAGAGATTCTTCGACTGGCGCTCGGCAGCGGCGCTGTTGATGACCAACTCTCTTGT  20836

Qy   2754  CGTTCATTATTCGACTGAGGTAAGCGCCTCACTTCACCATTATTATATACCATCTCTTGT  2813
           |||||||||||||| | || ||||| |||||| |||||||||| || |||||||||||
Db  20837  CGTCCTTCTTCCACTGAGGTAAGCCACCTCCACCATTGTAATATACCATCTCCTG-  20895

Qy   2814  GATTTGTTCTTTATTTTGTTCAGTGATTGTCCTGAGTACTCTCTTCCAAGTT  2873
           ||||||||||| ||||||| |||| ||  | ||| |||||||||||||| |
Db  20896  ------TTTTTATTTGTTTGTTCGTGATGTTTTTGTCCTGAGTACTCTTCCAAATT  20948

Qy   2874  ATTCACACAGTGACTGTGTGATTTAAGTTGGGGGAGGGCTCAGGA--GTATGTTGCA  2931
           |||||||||| |||||||| ||||||||||| ||||||||||||  ||||||||||
Db  20949  GGTCACACAGTGACTGTGTGATTTAAGTTGGGGGAGGGCTCAGGAAGTGTGTTGCG  21008

Qy   2932  TTGTATATATTTTAAGTCTGCATTCATCTAAGGCATAGAAAACC----AAAAAAAAATT  2988
           ||||||||||||||| |||||||||||| |||||| |||| || ||    |||| ||||||
Db  21009  TTGTATATAATCTTGAGTCTGATCTGATTCCACCATTGGCATATAGGGGATAATGATGAGA  21068

Qy   2989  AAAAATTTCAGAAAATGATTTCACAAAAAAAAAGAGTGTTCATTACATTTAG  3048
           ||||||| |||| |||||||| |||||||| | || || |||||| |||||
Db  21069  AAAAATTTCAGAAAATGATTTCACAAAAAACAGAGTGTTCATTAGAATGTGAGA  21128

Qy   3049  GATCAAGTCTAGAGTCTTCATTGGAGTTGCATATGCATAGGGATAATGATGAGA  3108
           |||||||||||| |||||||| ||||||||||||||| |||||| ||||||||
Db  21129  GATCGAGTCTAGAGTCTTCATTGGATCATCAGTTATATAGCATATGCATAGGGACACTATGAGA  21188

Qy   3109  TAGCCTTGTAAGCA  3122
           ||||||||||||||
Db  21189  TAGCCTTGTAAGCA  21202
```

```
Db  19158  GAGACGAAGCTGTTGAACAGGCCTAAATGAAGTCCCAGTGGAGGAGTACATCAGACTTTTC  19217

Qy   1207  GAGATGAATGACTTCTGGGAACGAGGTATCCCTGATAATGAGACTTTAGCCCAGTTGGGG  1266
           ||||||||  ||||||||| ||||  | ||| ||||| || ||||||||||||||| ||
Db  19218  GAGCTGAACGACTTCTGGGGAACGCAGCGAACGAAGTACCCCTATTATCAGACTTTAGCCCAGCTGGGG  19277

Qy   1267  TTACTGGAGGACGTGCAGCATCTGTTCGAGAGTGTCATCTG----------  1308
           ||||||||||||||  ||| | || | ||||| | | || 
Db  19278  CTACCGGAGGACGACCTACACACCTATCGAGAGTGTCATCTTAC  19337

Qy   1309  ----------ATAAGGAGGAGCAATCGAGTTGTTCTTTCCACACTGCAAGTGGAAATG  1355
                     |||  ||| ||||| ||| |||||||| ||| | || |||| |
Db  19338  CCGTACGTCGCTTACAAGAAGAAATAAGAGAGTTTCTCTCCACTCGTGCAAGTGAGATG  19397

Qy   1356  TATGAGGGACTCACAGACTTTGAGCTGATACCATGGGTTAGGCTTCTTGACGTTCTTA  1415
           |||| |||||| ||||||||||||||||| || || ||| | |||||||| |||||
Db  19398  TATCAGGGACTTACAGCAGATGAGCTGAGAGTGAAGTGGGGTTCTTGACTTTTTCC  19457

Qy   1416  GTGGATGAACAGCGGTACCAGCATTTAGATCAAGAAATTGGAAGAACTGTTTGGTTTCCCT  1475
           |||||||||||| || |||| | | || |||| ||| |||| || | || ||  |||||
Db  19458  GTGAGCGAGCAGTGTTACCAGTCATCATCAGGGATGTTGGAGGATTGTTTTGGCTTACCC  19517

Qy   1476  AGTGGAAAGGGAACCAACCCAGGTTGACAGGGAGCTTAAGGAATTGTGGGCTACT  1535
           || | || || | |||||||| ||| || || ||| ||||| ||||||| ||||
Db  19518  AGTGGAAAGGGAACTAAACCCAAGTTAGACGAGGGAGAGTTGAAGGATTTGTGGTTAACT  19577

Qy   1536  ATTGGGAACAATCTACCGCTAAACTCGACGCGGTCCAAGACGAACCAAAATCCGAGTCCT  1595
           |||||||| || |||| || ||||||| | |||| ||| ||| ||||| ||| |||
Db  19578  ATTGGGAACTAACGCTGCGCTCAACTCTGCGACGTCTATGGCAGATTCGAAGCCT  19637

Qy   1596  GTGATTCGCTACTTTCAGCGCTGCGGTTGCCAATGTTTTTACTCACGGGAGTCTACAGGC  1655
           ||||||| | |||| |||| | | |||  | ||||| |||| |||| || ||||||||||
Db  19638  GTGATCTGCTACTATCAACGCTCTATAGCCGAATGTTCTGTACTCTAGGGAATTCCGTAGA  19697

Qy   1656  ACCGTGTCTACAACAGACGATCGATGAGATGATGATTCAGCGCGCTTATAGGGATTCTCCGCCTT  1715
           ||||||| || ||| | || || ||| ||||||| ||||| | |  ||| || ||||| || |||
Db  19698  ACCGTGTCTAACAGAGACAGATGATGAGATGATGATTAATTCGCACTTCACTTCACCAGTAATGCCTCTG  19757

Qy   1716  ACAAAAGAAAGAATGTCCTGAGAGATCTTAACGACTTACAGACTTGCGCCGATGCGTTGT  1775
           ||||| |||||| ||| ||||||| | |||| ||||| || ||||| || || |||
Db  19758  ACAAAGGGAGGAAGTCCCCAGGTTTGATCCGAGGATGATGATGCAGCCAGTTATGCTTCTG  19817

Qy   1776  TTGATCCATCTGTGTGGGTACATGAAGGGCGCTGACAAACGGCAAGAGAAGAAGTAAGA  1835
           |||||||| ||||||  |||| || ||||| | | ||||||| ||||||||||| ||
Db  19818  TTGATCCACCTGTGTGGATACAGAGGAGTGGGGGCCACCACCAGGAGAAGAGAGGGCGA  19877

Qy   1836  GGAGCACTATGCGTGCGTTGGTGCCGTTGTGACCGCCAATTCTGAAAGTTTGTGGAGGTTTCCGCTC  1895
           ||||  | | ||||||| | || | | |||| ||| ||||||||| ||| | ||| ||||
Db  19878  GGAGCCCTTGTGTAGGGGTGGTGTGTTGTGACACCACTGATTACATGTGTGTAACCTTTC  19937

Qy   1896  AAGGAGTAGGGTTAGCACCGAGAATGATGACTTGGATCATCCTTGGCCCGATGTGAGTTC  1955
           |||||||| | |||||||||||| |||| |||| || |||||| |||| || |||||| |
Db  19938  ACGTCTCCAGGGTTTGATCCGGAGGATGATGATGAGTCATCTTGGCGTCGTTGTGTGAGTTT  19997

Qy   1956  TCTGAGTTTGACATGGTTGGCGACTTCACCGCTACAGGTTCGAGCATTCATCGATTAGA  2015
           |||||||| |||| | || ||| || ||| || | ||| ||| | || ||||| |||
Db  19998  ACGTCTGGCAGCACATGGTTGCGATTTCTATCGCTACTACTTGAGCAGCCGTGATCTGACCCGA  20057

Qy   2016  ATCGCCAACATTCTTTCCCCTGCATTAGTTTATTCGAGGGCCAGTCCGCCAACTGAGGAGATT  2075
           || || ||||| | || ||||| ||| ||  || ||| | |  ||| ||||| ||| |||
Db  20058  CTGGAGCACGACATTGTGCTTCTGCTATGCTCTACATTCGAGCAGCAGAGTAGGGATAGGGAAACATT  20117

Qy   2076  GACTTCAAGGCCTGCGCTTGAAGATCTTTATTTCGGAGTTCCGCCCAACTGAGGAGATT  2135
           |||||||||| || ||||| || ||  |  || ||| | | |||| ||| |||||||
Db  20118  GACTTCAAGGCCTGCGCTTGCGCGTGATTGCTGTACTCTGAGAGCCGTCCACCGACGATGACA-A  20176

Qy   2136  AGTCACACCGAAGGAGCTGAGTATGCGCAGCATGTATCATTCAGTGAGCATGAGATCAGGCGC  2195
           || ||||| || || ||| || || | ||||||| |||  |||||| | | ||| || | |
Db  20177  CGTCCCTACAAAGAAGCTACCCAAGCTGAGATTGCTGAGACAGATAGCAGAGAGAAAGCAAG  20236

Qy   2196  GAGTTTGACACGAGCATGTATCATTTCAGTGAGCATGAGATCAGG  2255
           ||||||||||||||| |||||| |||||||||| || ||||||||
Db  20237  GAGTACGATACGCGCCATGTGAGCATGTACCTCCCAGCGAGCCGGCGAGCAAG  20296
```

XX
DE   BAC containing repeats from centromeres 1-4 #2.
XX
KW   Centromere; microsome; vector; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200055325-A2.
XX
PD   21-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-US007392.
XX
PR   18-MAR-1999;   99US-0125219P.
PR   01-APR-1999;   99US-0127409P.
PR   18-MAY-1999;   99US-0134770P.
PR   13-SEP-1999;   99US-0153584P.
PR   17-SEP-1999;   99US-0154603P.
PR   16-DEC-1999;   99US-0172493P.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Preuss D, Copenhaver G, Keith K;
XX
DR   WPI; 2000-587529/55.
XX
PT   Recombinant DNA construct comprising a plant centromere, useful for
PT   producing stably inherited microsomes which can serve as vectors for the
PT   construction of transgenic plant and animal cells.
XX
PS   Claim 102; Page 321-335; 1449pp; English.
XX
CC   The present invention relates to a recombinant DNA construct of a plant
CC   (Arabidopsis thaliana) centromere. The constructs are useful for
CC   producing stably inherited microsomes which can serve as vectors for the
CC   construction of transgenic plant and animal cells expressing selected
CC   proteins such as hormones, enzymes, interleukins, clotting factors,
CC   cytokines, antibodies, and growth factors
XX
SQ   Sequence 64415 BP; 18698 A; 13554 C; 13083 G; 18980 T; 0 U; 100 Other;

Query Match   42.9%; Score 1339; DB 3; Length 64415;
Best Local Similarity   75.2%; Pred. No. 0;
Matches 1812; Conservative 0; Mismatches 490; Indels 109; Gaps 7;

```
Qy    788 ACTCAACCGGTTACTCGACCGTACGCTGGTTGAGTTCGTTCGTTTGTTTTGCTATTA 847
              |||| |  |||  |||| ||  ||  |||||||| |||| |||||| |||
Db   2012 ACTCGACCGTATTACTCGA-CTGCGTACTTGACCGGTTTAGTGTTTAGTGTTTATTTAGA 2070

Qy    848 CTTAACATATTAAACGTTTATCTTTGAGTTTCGTCTGTTTTAGGTT-TCATCATGAGTAAC 906
              |||||||||| || |||  ||||| ||||||
Db   2071 CTTAACATATTGATATTTGGCTTTGAGTTTGAGTTACATTCTTTTCAGGAAATTAATATGAGTAAC 2130

Qy    907 TACAGTGGAAAATCCTCTATGGACCCTGATTATAATGTGGATGAAGCTAAGTGCCTGTCC 966
              |||||||  |||||||| || |||| ||||| |||| |||||||| ||| ||||| |
Db   2131 TACAGTGGCGAATCCTCCATGATGCGGATTACAACGTGCTAGAGAGCTGATGAAGCCTGTCA 2190

Qy    967 ACTAGACCGGAGTGAGAGCTGAGAGAGAGATATCAAAAACTGAGTTGAAATTTGAAACGCTCT 1026
              |||| || |||| ||||  |||| ||  ||||  | |||| ||
Db   2191 ACTGGACCAGAGAGAGAGAGAGAGAGAGGAGCTGCGGATTACAACGTTCAGAGGAGCTTATGAGCTTA 2250

Qy   1027 GCAGTCGACGTAATCAACAATGGCAGTGAAATCGGTTAGAAATCCTGATATGTTGAGT 1086
              ||||||| ||||||||  ||||  |||| ||||||||||| |||||||||
Db   2251 GTAGCTCACGCAATGGCTGCTGATGAGAGAGTATCAAAAACTGAGTTGATGAGGCGCAAG 2310

Qy   1087 AGATATGAGCTGATTGATGAGATATCAAAAACTGAGTTGAAATTTGAAACGCTCT 1146
              |||| ||||||  |||||||||||||||||||| |||||| ||  ||||||| |||
Db   2311 AGATATGAGCTACGACGAGATATTGACGTCGAGTGGACTGGCCTCAGAGGCCTCACAGA 2370

Qy   1147 GAGACGAAGCTACTGAACATCCGACACAAATCCGAGGTTACAGTTGACAGGAGTATATCAGATTCTT 1206
              ||||||||  |||||||| |||| ||| ||||||| || ||||| ||| |||||||  ||
Db   2371 GAGACGAAGCTCGTTGAACAGGAGGCCTGATGATGCGCGGGAGAGCGGCAACAACAGCAGGAT 2430
```

```
Qy   1207 GAGATGAATGACTTCTGGGAACGAGGTATCCTGATATGAGACTTTAGCCCAGTTGGGG 1266
              ||||||||  ||||||||||||||||||  ||||||||| |||| |||||||||||
Db   2431 GAGCTCAATGACTTCTGGGAACGAGGTACCCTGTTATCAGACTTTGGCCCAGTTGGGG 2490

Qy   1267 TTACTGGAGACGTCCAGCATCTGTTCGAGAAGTGTCATCTG--------- 1308
              |||||||||||| |  |  |||||||||||||||||||
Db   2491 CTACTGGAGACGTACAACACTTATTCGAGAAGTGTCATCTGGAGACGCTGATGTCTTAC 2550

Qy   1309 ---------ATAAGGAGGAGCAACAATGAGTTCTTCCACACTGCAAGTGGAAATG 1355
                       |||| || ||||||||||   |||||||||||  |||||| ||||
Db   2551 CCATACGTTGCTTACAAGAAGCAACAATAGAGTTTCTCTCACTCTGCAAGTGGAGATG 2610

Qy   1356 TATGAGGGACTCACAGACTTTGAGCTGGATACCATGGGGTTAGGCTTCTTGACGTTCTTA 1415
              |||  |||| || |||   |||||||| || || ||  ||  ||||||||| ||| ||
Db   2611 TATCAGGACTTACAGCAGATGAGCTGGGAGAGTGAGGGGTTGGGGGGTTCTTGACTTTTTCA 2670

Qy   1416 GTGGATGAACAGCGGTACCAGTTAGATCAAGAAATTGAAGAACTGTTGGTTTCCCT 1475
              ||| ||||||||||| ||||| | | ||||||  ||||| |   ||||| |||||||
Db   2671 GTGAACGAGCAGCGTTACCAGCTATCTCATCAAGAGCTTGAAAGGATTGTTTGGTTTCCCC 2730

Qy   1476 AGTGGAAAGGGAACCACCCAGGTTTGACAGGGAAGAGCTTAAGGATTGTTGGGCTACT 1535
              | ||||||||||||   |||| |||||   ||||||||  ||| |||||  || ||
Db   2731 AATGGAAAGGGAACTAAACCCAAGTTCGAGAGGGAAGAGAGTTGAAAAATTTGTGTTAACC 2790

Qy   1536 ATTGGGAACAATCTACCGGTCAAACTCGACGCGGTCCAAGACGCAACCAAATCCGGAGTCCT 1595
              |||||||||  |  || |  |||| ||  ||| || ||||| |||||||||| |||| |||
Db   2791 ATTGGGAACGAGATATAGGCTCAACTCTGCGAGGTCTAAGAGCAACCAGATTCGAAGCCCCT 2850

Qy   1596 GTGATTCGCTACTTTCAGCGCTCGGTTGCCAATGTTTTTACTCCAGGAGTCTACAGGC 1655
                              |||| |  ||||  ||||| |||||| ||| |||||||||
Db   2851 ----------------GTAGCGAATGTTCTGTACTCTAGGGAATCTACAGGC 2886

Qy   1656 ACCGTGTCTAACACAGACATGAAGATGATAGATTCAGCGCTTATAGGGATTCTCCGCCTT 1715
              | |||||||||||||||||||||||||||| ||||  || | | |||| ||||||||
Db   2887 ATCGTGTCTAACACAGACATGAAGATGATAGATTCGCACTCTCCAGGGCATTCTCCGCGAGA 2946

Qy   1716 ACAAAGGAAAGAATGTCCTGAGAGAGATCTTAACGACTCACCACCAGTAATGCCTCTG 1775
              ||||| || |||| |||||||||  |  || | ||  |||||||||| |||| |||
Db   2947 ACAAAGGGAAGAAGGTCCTGAGAGGGCAACCTCAATGCATGTCTAAGGGCCGGTTATGCTTCTG 3006

Qy   1776 TTGATCCATCTGTGTGGTACATGGTGGTACTGAAGTGGGCGCCTGACAAACGCAAGAAGAGGTAAGA 1835
              |||||| ||||||||| | |   |||| ||| || ||| ||| ||||| |||||||| ||||
Db   3007 TTGATCCACCTGTGTGGATACAGGAGGAGGTGGGCCGCCACACCAAGCAGAGAAGAGGTGCGA 3066

Qy   1836 GGAGCACTATGCGTGGGGTGGCGTTGTGACGCCAATTCGAAAGTTTGTGGAGTTCCGCTC 1895
              |||| ||| |  |  ||||| ||| |||| ||  || || |   ||||| || |||||
Db   3067 GGAGCCCTTTGTGTAGGTGGTGGTGTGTGACCATTCTGATTGCATGGTGGTACCGCTC 3126

Qy   1896 AAGGAAGTAGGGTTAGCACCGAGAATGAGATGAGTCTTGCGCCGCCGATGTGAGTC 1955
              |    |   ||||| || ||||| || |||| |  || |  ||  ||| |||||
Db   3127 ATGTCTCCAGGGTTTGATCCGAGATGATGAGGTTAGATCATTTCGCGTCGTTGTGAGTTT 3186

Qy   1956 TCTGAGTTTGACATGGTTGGCGGACTTTCACCCGTGAAGGTCATGAGTTCATCATGATTAGA 2015
              | || | |||||||||||| | ||||||| |  | |||   ||||| ||||| |||||
Db   3187 CTTGGAGCACGACATGGTTGGCGATTTTCATCGCTACAAATTCGACTACAAATTCGACACTCCGA 3246

Qy   2016 ATCGCCAACATTCTTTCCCTGCATTTACGCTACTAGGATTCTGAGCGGCAGGAACATT 2075
              |  |||||||| |  ||||| | ||  |||| | |  |||| | |||| ||||||||||
Db   3247 ACAGCCAACATTCCTGCTTCCCTGCAGTTCACTCTACTACTTCAGGGGCGGAAAACATT 3306

Qy   2076 GACTTCAAGCCTGCGCTTTTATTTCGAGGAGTCTTATTTCGAGGAGTGGGGGCGGAGATT 2135
              ||||||| ||||||||     | |||  | |   ||| |||||| |||||||| ||||
Db   3307 GACTTCAATCCTGCGCGTGATTACCTCGTGTATTCCTCTACTTTGGAGAGTCTCGAGAAGC 3366

Qy   2136 AGTCACACCGAAGGACTCAAGATGGAAACTCGAGTAGAGGGCTCGAAGAGCAATGACCAGC 2195
              ||  ||   |||| | || | |||||| |||||  ||||   |||||||||||||||||
Db   3367 GGCCCTACAAAAGAGCTACAGAGGATGAGATTGCTGAGACAGATAGGATAGAAAGGAG 3426

Qy   2196 GAGTTTGACACGAGCATGTATCATTTCAGTGACGAGTATATCCCAGCGGAAAAGCAAG 2255
              ||||||||| |||| ||||| ||| |||||| |||||||||  |||| ||| ||||||
Db   3427 GAGTACGATACGAGCATGTATCATCATTTCAGTGACGAGTACCTCCAGCGCGGGAGCAAG 3486

Qy   2256 AGTTTGAGCGAAGCTCAGACAGGACAGGCCTGATGAAGTTACGAGGAACAACAGCAGGAT 2315
```

```
XX
KW   Centromere; microsome; vector; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200055325-A2.
XX
PD   21-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-US007392.
XX
PR   18-MAR-1999;   99US-0125219P.
PR   01-APR-1999;   99US-0127409P.
PR   18-MAY-1999;   99US-0134770P.
PR   13-SEP-1999;   99US-0153584P.
PR   17-SEP-1999;   99US-0154603P.
PR   16-DEC-1999;   99US-0172493P.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Preuss D, Copenhaver G, Keith K;
XX
DR   WPI; 2000-587529/55.
XX
PT   Recombinant DNA construct comprising a plant centromere, useful for
PT   producing stably inherited microsomes which can serve as vectors for the
PT   construction of transgenic plant and animal cells.
XX
PS   Claim 102; Page 386-404; 1449pp; English.
XX
CC   The present invention relates to a recombinant DNA construct of a plant
CC   (Arabidopsis thaliana) centromere. The constructs are useful for
CC   producing stably inherited microsomes which can serve as vectors for the
CC   construction of transgenic plant and animal cells expressing selected
CC   proteins such as hormones, enzymes, interleukins, clotting factors,
CC   cytokines, antibodies, and growth factors
XX
SQ   Sequence 83390 BP; 24664 A; 17305 C; 17224 G; 24140 T; 0 U; 57 Other;

Query Match        42.0%;  Score 1311;  Pred. No. 0;  DB 3;  Length 83390;
Best Local Similarity   75.2%;
Matches 1809;  Conservative    0;  Mismatches  470;  Indels  126;  Gaps   9;

Qy    795 CGGTTACTCGACCGGTACGCTGGTTAGATGTGTTTTTGGTTTGCTATTACTAACAT 854
          |||||  |||||||||||||   ||||| ||   |||||||||| ||||||| ||||
Db  23635 CCGGTTACTCGACCGGTGAACTTGATT--GGTTTAGTGTTGTTTATTTGGACTAACAT 23578

Qy    855 ATTAAGTTATCTTGAGTTCGTCGTTTTAGG-TTTCATCATGAGTAACTACAGTG 913
          ||| | ||| |||||| |  |||   |||||||||  ||||||||||||||||||
Db  23577 ATTGATATTTGTTTTAAGTTACATTATTTTTCAGGAAATCAATATGAGTAACTACAGTG 23518

Qy    914 GAAAATCCTCTATGGACCCTGATTATAATGTGAGACCTAAGTCCTGGTCCACTAGAC 973
          | ||||||||| |||| |  |  |  |||  ||   ||||||| ||||| |||| ||||
Db  23517 GCGAATCCTCCATGGATGCAACGTCGATGATGAAGCTGATGAAGCTGGTCAACTCGGAC 23458

Qy    974 CGGAGTGAGAGCAACATGTTTACGAGAGCTATAGGAGTGAATTGAACGCTCGCAGCTC 1033
          | ||| ||| || || ||  ||||||||| | ||||| | ||   |||||||||||||
Db  23457 CAGAGAGAGCGACACAGGCTTATGAGAGCTTGAGAGTGCTGAGA--GAAAGAGAGCAATGA 23398

Qy   1034 GACGTAATCAAAGAGAGCTGAAATCGCAGGATGATGCGCGATGTCGAGTAGATATG 1093
          ||||  || ||||||||||||| | |||||  || ||||| || ||||||||||||
Db  23397 GACGCAATGAGAGAGAGCTGAGATTGCTAGA-GAAAGAGAGCAATGACCAGCAGATATG 23339

Qy   1094 AGCTTGATTGATGAGGATATCAAAACTGAGTATGAGCCGCAAGGAGGACGA 1153
          || ||||| || ||| |||||   |  |||| |||| || |||||| |||||
Db  23338 AGTTGATCGACGAAGATATTGACGTCGAGGTTACGTGGAGGAGTATATCTTTGAGATGA 23279

Qy   1154 AGCTACTGAACAAATCCGACGAGGTTACAGTGGAGGAGTAATCAATATGAGTAACTACAGTG 1213
          ||||  |||||||| ||                    |||| |||||||||| ||||||||||
Db  23278 AGCTGTTGAACAAGCCT--------------------GATGA 23257

Qy   1214 ATGACTTCTGGGGACGAGGTATCCCTGATATGAGACTTTAGCCAGTTGGGGTTACTGG 1273
          ||||||||||||  ||||||||||| ||||||||||||||||||| ||||| ||||||
```

```
Db   3487 ...AGCTTGAGCGAAGCTCACAGAACAACAGTAAGTTGCAGAGGTGGTGCAAGAAACAAGAT 3546
             |||||||||||||||||||||||||||||||||||
Qy   2316 AAGTTACTGCCAAGTGCCTCAGGGCTATCAAGTTCTGAAGGACAAGATCAGCTGCTCC 2375
             ||||||||||||| |||||||
Db   3547 AGGCTACTCATCAAGTGCTTCAAGGCCATCAAGTTCTGACACAAGCTAAGCTGCTCC 3606

Qy   2376 TCTTCCACTACAACTATTCCGCAATGACAGCTCCCTCAGGACATGACATGCCTTCGAGGAGATAT 2435
             |||||  |||||||||||||||||  || | | |  |||||||||  |||||| ||||||||||
Db   3607 TCTTCTACCACCAGCTATTCCGCAGGGAGAGCCTCCTCCTCCAGGACATGCCCTCGAAGAGATAT 3666

Qy   2436 GACGC---------------------GCCCGAGCCTAGA 2453
             |||||                     |||||
Db   3667 GACGCGCCAGAGCCAACTCGCCACTGGCCTGAGCTAAGTCACCACCAGCCTGAGCCTTAGT 3726

Qy   2454 GAGCAGAAGATTCGATGTCCCGCGAGGCATTCATCATTCGAGCCTTCGTTGAATCTTAGG 2513
             |  | | || |||||| |||| | | ||||||||||||  |||| |    || |||||||
Db   3727 GACCGAGTAGTCCACCAGTCCCTGTGTGGCATTCATCATTCAAGCCTCGGGAGCTCGGG 3786

Qy   2514 AAGAAATAGGAGACGACACTCACTCGATCTAGCAGCGAGCAGAGACGACACTTCTGCAGTCT 2573
             | ||| || ||| || ||||||||||| ||||||| |||||| || ||||||||| |||||
Db   3787 AGAAAGAAGAGGCTCACCTGCTCACTCGGTCTCGGCAGTAGGAGTACACGACTTCTCCAGTCC 3846

Qy   2574 CGTAGTTTACGCGACCCGGTTGCTCGGCCCAATAGAAGAAGAGAGGTCGAGTATCCTCAG 2633
             ||||| |||||||||   |  | | || |  |||||||| |||| |||||||||| |||
Db   3847 CGTAGCTTACGCGACCAACCACGGTGATGGCCGCCAGCATAAGAACAGAGGTCGAGTATCATCAG 3906

Qy   2634 AGCGGTGCTGCCGCCACAGAGCTGATGAGATCGAGTACCCACATGCTGAGCTGATACG 2693
             ||||| ||||| ||||  || | |  |  |||||||||||  |  | | |||||| |||
Db   3907 AGCGGCTGCTGGCCGCCGACGAAGGAGCAGAGGTCGAGTACCCCCAGGGGGAAGCTAAGACA 3966

Qy   2694 GAACATGGCGGTTCGTCTATGGCTTGGGAGCAATCACAGCAGCCATTGACTACCAACTT 2753
             ||||  || | ||| |||||||| |||||||||||||| ||||| ||||| |||||||
Db   3967 CAACAGGAGGATTCTTCGATGGCCTGGGAGCAATCACAAGCAGCTATTGACGACCAACTC 4026

Qy   2754 CGTTCATTATTCGACTGAGGTAAGCGCCTCACTTCCACCATTATATATCATCTCTTGT 2813
             || || ||||||||||||||||||| ||| | |||||||||| |||| |||||| ||
Db   4027 CGCTCCTTCTTCCACTGAGGTAAGCACCTATCTCCACCATTGTAATATACCATCTCCTG- 4085

Qy   2814 GATTTGTTCTTTATTTGTTTCAGTGATTGATTTGTCCTGAGTACTCTCTTCCCAAGTTT 2873
             | ||||  ||||||||||||  ||| |   |||||||||||||||||||||| |||||
Db   4086 ------TTTTTATTTGTTTGTGTGATGTGTTTTGTCCTGAGTACTCTCTTCCAAATTT 4138

Qy   2874 ATTCACACAGTGGACTGTGTGATTAAGTTTGGGGGAGGGCTCAGGAA--GTATGTTGCA 2931
               ||||||||||||||||||||||| |||||||||||||||||||||  || |||||||
Db   4139 GGTCACACAGTGGACTGTGTGATTTAAGTTTGGGGGAGGGCTCAGGAAGTGTGTGTTGCA 4198

Qy   2932 TTGTATATATTTTTAAGTCTGCATTCATCTAAGGCATAGAAAACCAAAAAAATTAAA 2991
             |||||||||| ||  |||||||||||||||| ||||||||||   ||||||||||| |
Db   4199 TTGTATATATCTTGAGTCTGCATTCATCTGAGGCATAGAAAAAAACCCAAAAAAATTGAAA 4258

Qy   2992 AATTTCAGAAAATGATTTCACAAAAAAGAGTGTTCATGTAGTTGCATTACATTTAGGAT 3051
             ||||||||| |||||||||||||||| |||||||||||||||||||| ||||||||||
Db   4259 AATTTCAGAACATGATTTCACAAAAATAGAGTGTTCATGTAGTTGCATTGCATTTAGGAT 4318

Qy   3052 CAAGTCTAGAGTGTTCATTTAGGATTGTTGCATATGCATAGGGATAATGATGAGATAG 3111
             | |||||||||||| | |||||||||||||||||||||||||||||||||||||||||
Db   4319 CGAGTCTAGAGTGTTTCGTTTAGGATTGTTGCATATGCATAGGGATAATGATGAGATAG 4378

Qy   3112 CCTTGTAAGCA 3122
              ||| ||||||
Db   4379 TCTTTGTAAGCA 4389
```

RESULT 10
AAF22283/c
ID   AAF22283 standard; DNA; 83390 BP.
XX
AC   AAF22283;
XX
DT   20-MAR-2001 (first entry)
XX
DE   BAC containing repeats from centromeres 1-4 #6.

```
Db  23256 ACGACTTCTGGGGAACGAGGTACCCCTGTTATGAGACTTTAGCATAGTGGGCTACTGG 23197
Qy   1274 AGGACGTGCAGCATCGTTGTTCGAGAAGTGTCATCTG------               1308

Db  23196 AGGACGTACCATCACTTATTCGAGAAGTGCCATCTGAGCGCTGATGTCTTACCCGTACG 23137
Qy   1309 -----ATAAGGAGGAGGAACAATAGAGTTTCTTTCCACACTGCAAGTGGAAATGTATGAGG 1362

Db  23136 TCGCTTACAAGAAGGAACAATAATAGAGTTTCTCTTCACTCTGCAAGTGGAGATGTATCAGG 23077
Qy   1363 GACTCACAGACTTTGAGCTGGATACCATGGGGTTAGGCTTCTTGACGTTCTTAGTGGATG 1422

Db  23076 GACTTACAGCAGAGCTGGAGAGTGGAGAGGGTTGGGGTTCTTGACTTTTTCAGTGAACG 23017
Qy   1423 AACAGCGGTACCAGATTTAGATCAAGAAATTGGAAGAAGACTGTTTGTTTCCTTAGTGGAA 1482

Db  23016 AGCAGCGGTTACCAGCTATCTATCAAGAGCTTGGAAGGATTATTTGGTTTTCCCAGTGGAA 22957
Qy   1483 AGGGAACCACCCCAGGTTTGACAGGGAAGAGCTTAAGGATTTGTGGGCTACTATTGGGA 1542

Db  22956 AGGGAACTAAACCCAAGTTCGAGAGCGGAAGAGTTGAAGGATTTGTGGTTAACCATTGGGA 22897
Qy   1543 ACAATCTACCCGCTAAACTGCGACGCGGTCCAAGAGCAACCAAATCCGGAGTCCTGTGATTC 1602

Db  22896 ACGATTTGGCGTCAACTTTGCAAGGTCTAAGAGCAACCAGATTCGTAGCCCTGTGATCC 22837
Qy   1603 GCTACTTTCAGCGCTTCGGTTGCCAATGTTTTTACTCCAGGAGTCTACAGGCACCGTGT 1662

Db  22836 GCTACTATCAGCGCTCAGTAGCGAATGTTCTGTACCCCAGGAATCTACAGGCACCGTGT 22777
Qy   1663 CTAACAGACACATGAAGATGAATAGATTCAGCGCTTATAGGGATTCTCCGCCTTACAAAAG 1722

Db  22776 TTAACACAGACATGGAGATGATTGATTCTCCACTCAAGGCCATTCTCCGGAGAACAAAAG 22717
Qy   1723 GAAAGAATGTCCTGAGAGAGATCTTAACACTCACCACCAGTAATGCCTCGTGTTGATCC 1782

Db  22716 GGAGAGAAGGTCCTAAAGGGCGACCTTAATGATGCACCGTTATGCTTCTGTCGTTGATCC 22657
Qy   1783 ATCTGTGTGGGTACATGGAAGTGGGCGGCGCTGACAAACGGACAAGAGAGGTAAGAGGCAC 1842

Db  22656 ACCTATGTGGATACAGGAAGTGGGCGTGGGCGCCACACCAACGGAAGAGAGAAGTGCCGAGAGCCC 22597
Qy   1843 TATGCCTGGGGTTGTGACGCCAATTCTGAAAGTTGTGGAGTTCCGCTCAAGGAAG 1902

Db  22596 TTTGTGTAGGTGGCGTTGTGACCGATTCTGATTCGATGTGGCATGTGGTACCTCTCACGTCTC 22537
Qy   1903 TAGGGTTAGCACCGAGAATGACGCTTGATGGACTTGATCACTTGCGCCGATGAGTGTCTCTGAGT 1962

Db  22536 CAAGGTTGATCCGAGGATGATCGGAGGATGATTTAGATCACTTGCCGTGTGTTGTAGTTTCTGGAGT 22477
Qy   1963 TTGACATGGTTGGCGACTTCACCGGACTTCACCGCTCATCTGACCGTTCAGAGTTAGAATGCCA 2022

Db  22476 ATGATAATGTTGGCGGATTTTCATCGTCTACAAATTCAGCCACTCCCTGATCCGAACAGCCA 22417
Qy   2023 ACATTCTTTTCCCCTGCATTTACGCTACTAGGATTCTCGAGGGCAGGAACATTGACTTCA 2082

Db  22416 ACATTTTGCTTCCCGATCGAGGCCGATGGGCCTCCACCGACGTGATAATGTCCCTA 22357
Qy   2083 AGCCTCGCGTTGAAGATCTTATTTCGAGGCAGCAGTCGCCGCAACTGAGGAGATTAGTTCACA 2142

Db  22356 AACCTGCACGTGATTACCTCTACTTTGAGCGCCTCCACCGACTGATGATAATGTCCCTA 22297
Qy   2143 CCGAAGGAGCGAGTACAATAGAAGATGTTGATGAGACATATGATATAGATGAGGGCGAGTTTG 2202

Db  22296 CGACGGAAGCTATAGAGATGGATGATATTGCTGAGACGGATGAGGATGAGGAGGAGTACG 22237
Qy   2203 ACACGAGCATGTGATCATTTCAGTGAGACATATACCTCCGAGGAAAGACAGAGTTTGA 2262

Db  22236 ATACGAGCAATGTGTATCATTTCAGAGAGCACGTACCTCCGAGGCGCGGAGACAAGAGCTTGA 22177
Qy   2263 GCGAAGCTCACAGGAACAACAGCAAGCTGCAAGTGGTGCAAGAAAACAGGATAAGTTAC 2322

Db  22176 GTGAAGCTCACAGAAACAACAACAGTAAGTTCCGAGAGGTGTGCAGAGAAACAAGATAGGCTAC 22117
```

```
Qy   2323 TCGCCAAGTCGCCTCAGGCGCTATCATCAAGTTTCTGAAGGACAAGATCAGCTGCTCCTCTCCA 2382
Db  22116 TCATCAAGTGGCTTCAAAGCCATCACGTTTCTGACAGCAGCAAGCTTAGCTGCTTCTCTTCTA 22057

Qy   2383 CTACAACTATTCCCGCAATGACAGC--                                2440
Db  22056 CCACAGCTATTCCCGCAGGAGAGCGTCCTCAGGACATGCCTTCGAGGAGATATGACGCGC 21997

Qy   2441 ------------------------GCCCGAGCCTAGAGAGCAGA              2460
Db  21996 CAGGGCCAAGTCATCACAGGCCAAGTCACCACAAGCCTGAGCCTAGTGACCGAG 21937

Qy   2461 AGATTCTGCATGTCCCGCCAGGCATTCATCATTCGAGCTGCAATCAGGAGAATA 2520
Db  21936 TAGTCCCACCAGTCCCTGCAAGGCATTCATCATTCGAGCCTCGGGAGCCTCGGGGAGAAGA 21877

Qy   2521 GGAGAACGACACTCACTGATCTAGCAGCAGGAGCAGACAGCTTCTGCAGACGTCTCGTAGTT 2580
Db  21876 AGAAGGCTGCACTCGCTCGTCGTCGGCAGCAGGAGTACACGACTTCCCAGTCCCGTAGCT 21817

Qy   2581 TACGGGACCGCGGTGCTGCCGCAATAGAAGAGAGTCGAGTATCCTCAGAGCGGTG 2640
Db  21816 TACGGGCACCGGGGTGCTGGCCCAGCAGAGCAGGAGTCGAGTATCATCAGAGCGGTG 21757

Qy   2641 CTGGCCGCCGCCAGAGCTGATGAGATCGAGTACCCCATGCTGGAGGTGATACGGAACATG 2700
Db  21756 CTGGGCCGGCGACGAAGGACGCAGAGGTCGAGTACCCCAGTGGGGGAAGCTGAGACACAGCAGG 21697

Qy   2701 GCGGTTCGTCTATGGCTTGGGAGCAATCACAGGCAGCCATTGACTACCAACTTCGTTCAT 2760
Db  21696 GAGATTCTTCGATGGCGTCTATGGGGATTCTCCGCCTTATAGGGATTGTATTGTATA 21637

Qy   2761 TATTCGACTGAGGTAAGCGCCTCACTTCACCATTATTATTATATACATCTCTTGTGATTTGT 2820
Db  21636 TCTTCCACTGAGGTATGCACCTCACTCACCATTGTAATATACCATCTCTTG------ 21585

Qy   2821 TCTTTATTTTGTTTCAGTGATTGGATTGTCCTGAGTACTCCTTCCAAGTTTATTCACA 2880
Db  21584 TTTTTATTTTGTTTTGTGATGGTGTTTTGTCCTAAGTACTCTTCTTCCAAATTTGGTCACA 21525

Qy   2881 CAGTGGACTGTGTGATTTAAGTTTAAGTTGGGGGAGGGCTCAGGAAGTATA--TTGCATTGTATA 2938
Db  21524 CAGTGGACTGTGTGATTTAAGTTTAAGTTGGGGGAGGGCTCAGAAAGTGTATTGTATTGTATA 21465

Qy   2939 TATTTTTAAGTCTGCATTCATCTAAGGCATAGAAAA-ACCAAAAAAAATTAAAAATTTC 2997
Db  21464 TAATCTTGAGTCTGCATTCATCTAAGGCATAGAAAAAGAACCAAAAAAATTGAAAAATTC 21405

Qy   2998 AGAAAATGATTTCACAAAAAGAGTGTTCATGTAGTTCATTAGGATCAAGTC 3057
Db  21404 AGAAAATGATTTCACAAAAAATGAGTGTTCATTGTTCATTAGGATCGAGTC 21345

Qy   3058 TAGAGTGTTTCATTTAGGATGTGTTGCATATGCATATGAGATAGTGCCTTGT 3117
Db  21344 TAGAGTGTTTCGTTTAGGATTGTTGCATATGCATATGAGATAGTGCCTTGT 21285

Qy   3118 AAGCA 3122
Db  21284 AAGCA 21280
```

```
XX
OS   Arabidopsis thaliana.
XX
PN   WO200055325-A2.
XX
PD   21-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-US007392.
XX
PR   18-MAR-1999;   99US-0125219P.
PR   01-APR-1999;   99US-0127409P.
PR   18-MAY-1999;   99US-0134770P.
PR   13-SEP-1999;   99US-0153584P.
PR   17-SEP-1999;   99US-0154603P.
PR   16-DEC-1999;   99US-0172493P.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Preuss D, Copenhaver G, Keith K;
XX
DR   WPI; 2000-587529/55.
XX
PT   Recombinant DNA construct comprising a plant centromere, useful for
PT   producing stably inherited microsomes which can serve as vectors for the
PT   construction of transgenic plant and animal cells.
XX
PS   Claim 102; Page 529-549; 1449pp; English.
XX
CC   The present invention relates to a recombinant DNA construct of a plant
CC   (Arabidopsis thaliana) centromere. The constructs are useful for
CC   producing stably inherited microsomes which can serve as vectors for the
CC   construction of transgenic plant and animal cells expressing selected
CC   proteins such as hormones, enzymes, interleukins, clotting factors,
CC   cytokines, antibodies, and growth factors
XX
SQ   Sequence 90336 BP; 26524 A; 18837 C; 18625 G; 26350 T; 0 U; 0 Other;

Query Match            42.0%;  Score 1311;  DB 3;  Length 90336;
Best Local Similarity  75.2%;  Pred. No. 0;
Matches  1809;  Conservative    0;  Mismatches  470;  Indels  126;  Gaps  9;

Qy     795 CGGTTTACTCGACCGTACCGTGGTTAGATTGTGTTTTGGTTTGCTATTACTAACAT 854
           ||||| |||||||||| ||||||||||  | |||  ||||||||||| ||| ||||||
Db   36390 CCGGTTACTCGACCGTGAACTTGATT--GTTTAGTGTTTGTGTTTATTTGGACTAACAT 36333

Qy     855 ATTAACGTTTATCTTTGAGTTTCGTCGTTTTTAGG-TTTCATCATGAGTAACTACAGTG 913
           || ||  ||||| ||||||||||| ||  ||||||  ||||||||||| ||||||||||
Db   36332 ATTGATATTTGGTTTAAGTTACATTATTTTTCAGGAAAATCAATATGAGTAACTACAGTG 36273

Qy     914 GAAAATCCTCTATGACCCTGATTATAATAATGTGATGAAGCTAAGTCCTGGTCCACTAGAC 973
           ||||||||||||||| | ||||||||||||| |||||||| |||||||| |||||||||
Db   36272 GCGAATCCTCCATGATGCAGATTACAACGTCAGTATTACAACGTCGATGCTGAAGCTGAAGCTTGGTCAACTGGAC 36213

Qy     974 CGGAGTGAGAGCAACATGTTACGAGAGCTATAGAGAGCTTATGCGAGTTAAGAGAGTAGATATG 1033
           |||||||||||| |||||||||||| |||| ||||| || || || | |||||| |||||||
Db   36212 CAGAGAGAGAGGCGCAGGCTCATGATCAGAATGCAGATTACAACGTCGATGCAAGCTTAGGATGCGATGCAGTCGAGCTC 36153

Qy    1034 GACGTAATCAAAGAAGAGCTGAAATCGCTAGAGAGAAAGAGGGCGATGTCGAGTAGATATG 1093
           ||  ||  ||||  |||||||||| ||||| |||||| ||||| ||  |||||| |||||
Db   36152 GACGCAATCGAACGCGTGAGATTGCTGAGATTGCTGAGAAGAGAGGGCATGGATGTGTCGATGATGATGAGCAGATATG 36094

Qy    1094 AGCTGATTGATGAGGATATCAAAACTGAGTAGACGAGCCAGATGCGCCAGAGCGCCAGAGGAGACGA 1153
           |||||| ||| |||||||| |||| ||||  |  || |||||||| |||||||||||||||| |||
Db   36093 AGTTGATCGACGAAGATATTGACGTCGAGTATGCGGTGCAATGCCTGAGTCATGGCAACAGAGAAACGA 36034

Qy    1154 AGCTACTGAACAAATCCGACGAGGTTACAGTGGAGGAGTATATCAGATTCTTTGAGATGA 1213
           ||||| |||| || ||| ||||| | |||| |  ||  |||||||||||| ||||||||
Db   36033 AGCTGTTGAACAAGCT----------------------------------GATGA 36012

Qy    1214 ATGACTTCTGGGAACGAGCTATCCTGATATGAGACTTAGCCCAGTTGGGTTACTCGG 1273
           ||| ||||||||||||||||| ||||| || |||||| ||| ||||| ||||| ||
Db   36011 ACGACTTCTGGGAACGAGCGAGTGTACCCCTGTTATGAGACTTTAGCATACATAGCTGGGGGCGTACTCGG 35952
```

```
Qy    1274 AGGACTGCAGCATCTGTTCGAGAAGTGTCATCTG------ 1308
           ||||||||||||||| |||||||||||||||||
Db   35951 AGGACTGCAGCATCACTTATTCGAGAGAGTGCCATCTGGAAATGGATGCTGAGACGCTGATGTCTTACCCGTACG 35892

Qy    1309 ------ATAAGGAGGAGACAATCAGTTTCTTCCCACACTGCAAGTGCAAGTGGAAATGTATGAGG 1362
                 |||||||||||| ||||||||||||| |||||| || |  ||||||||||
Db   35891 TCGCTTACAAGAAGGAAACATAGAGTTTCTCTTCTCTCTGCACTCTGCAAGTGGAGTGTATCAGG 35832

Qy    1363 GACTCCACAGACTTTGAGCTGGATACCATGGGGTTAGGCTTCTTGACGTTCTTAGTGGATG 1422
           ||||| ||||||||||||| ||||| |  ||||| |||  |||||| || |||||| ||||
Db   35831 GACTTACAGCAGATGAGCTGGAGAGTGAGAGAAATTGGAAGAACTGTTTGGTTTCCCTAGTGAA 35772

Qy    1423 AACAGCGGTACCAGATTAGATCAAGAAATTGGAAGAACTTGGGAAGAGCTTAAGGATTGTGGGCTACTATTGGGA 1482
           | || |||||||||||| |||| || ||| || |||| ||| ||||| || || |||||  || ||| ||||||
Db   35771 AGCAGCGGTTACCAGCTCATCATCAAGAGCTTGGAAGGATTATTGGTTTTCCCAGTGGAA 35712

Qy    1483 AGGGAACCAACCCCAGGTTGACAGGGAAGAGAGCTTAAGGATTGTGGGCTACTATTGGGA 1542
           ||| ||||| |||||| |||| || ||||| | || |||| ||| ||| ||||  |||
Db   35711 AGGGAACTAAACCCCAAGTTCAGAGGGAAGAGAGTTGAAGGATTTGTGGTTAACCATTGGGA 35652

Qy    1543 ACAATCTACCGCTAAACTCGACGCGGTCCAAGACGAGCAACCAAATCCGGAGTCCTGTGATCC 1602
           || || |||||| || || ||| ||||| ||| |||| ||||| || || || |||||| |||
Db   35651 ACGATTTGGCGTCAACTTGCAAGGTCTAAGAGCAACCAGATTCGTAGCCCTGTGATCC 35592

Qy    1603 GCTACTTTCAGCGTCGGTTGCCAATGTTTTTACTCCAGGAGTCTACAGCCACCGTGT 1662
           ||||| ||||||| |||| |||||||||||||  ||| ||||||  ||||||||||||
Db   35591 GCTACTATCAGCGGTCGACGCTCAGTCAGTGGAATGTTCGTACCCCCAGGGAATCTACAGCCACCGTGT 35532

Qy    1663 CTAACACAGACATGAAGATGATAGATTCAGCGCTTATAGGGATTCTCCGCCTTACAAAAG 1722
           || || || | |  || |||| |||||  |||| ||  || | |||| || |||| ||||
Db   35531 TTAACACAGACAATGGAGATGATTGATTTCATCGCTCACAAATTCAAGGCATTCTCCGAGAACAAAGG 35472

Qy    1723 GAAAGAATGTCTGAGAGGAGATCTTAACGACTCACCACCAGTAATGCCTCGTTGATCC 1782
           |||||||| |||||| ||||||||| |||| |||||||||  |  ||| | |||||||
Db   35471 GGAAGAGGTCTTAAAGGGGGACCCTTAAAGCCCGGTTATGCTTCTTGTTGTTGATCC 35412

Qy    1783 ATCTGTGTGGTACATGAAGTGGCGCTGACAAACGACAAGAGAAGAAGTAAGAGGAGCAC 1842
           ||| |||  | || |  || | |  ||||||| |||||||| ||||| ||| |||||||
Db   35411 ACCTATGTGGATACAGGAGTGGGGCGCACACCAACGGGAAGAAGAAGAAGGTGCGAGGAGCCC 35352

Qy    1843 TATGCGTGGGTGGCGTTGTGACGCCAATTCTGAAAGTTGTGCTGAAGTTCCGCTTCAAGGAAG 1902
           || ||||  || ||||||| | |||||| || || ||||||  | || |||  ||| ||||||
Db   35351 TTTGTGTAGGTGGGCGTTGTGACACCGGATTCTGATTGCGATGTGTGTACCTCCACGTCTC 35292

Qy    1903 TAGGGTTAGCACCGAGAAATGATGAGACTTGCGCCGATGTGAGTTCTCTAGT 1962
           ||||||  |  | ||||| ||||  |  || |  | || | | || |||||
Db   35291 CAAGGTTTGATCCGAGGATGATGAGTTTAGATCGTTGTGTGTGAGT 35232

Qy    1963 TTGACATGGTTGGCGACTTTCACCGTCGGACTTCAGCGTTCAGCATTCATCATTAGAATCGGCCA 2022
           || || ||||||| ||| ||| | | || |||| || | | |||| | |  ||| ||||| |||
Db   35231 ATGATAATGGTTGGCGATTTTCATCGTCATCATCGAGCTGCTCAAATTGATCGATCCCGAACAGCCA 35172

Qy    2023 ACATTCTTTCCCCTGCATTTACGCTTACGGCGCCGCAGCATTGACTTCA 2082
           ||||| ||| || || |||| || || ||| |||| |||| |||||||
Db   35171 ACATTTTGCTTCCTGCCATTTGCTCGCTTGCGGCGGCCACCATACTTCAGGGTAGGAACATTGACTTCA 35112

Qy    2083 AGCTGCGCTTGAGAAGATCTTTATTTCGAGGTAGCTGAGGTAGTAGTCACA 2142
           |||| || ||||| |||||| ||||| || || |  | ||||| |||||
Db   35111 AACCTGCACGTGATTACCTCTACTTTGAGAGCGCTCCACCGACTGATGATAATGTCCCTA 35052

Qy    2143 CCGAAGGAGCTACAATGAGTGTTGATGAGACATATGAGGCGGAGTTTG 2202
           |||| ||||| || |  |||| | | | | |||||  || ||||  ||
Db   35051 CGACCGAAGCTATAGACGGCGTATGAGACGGATGATAGGGGAGGAGTACG 34992

Qy    2203 ACACGAGCCATGTATCATTTCAGTGACGAGCAAGCTGCAGCCTGAAAGCAAGGATGA 2262
           |||||| || | || ||||| || || ||  | || | |||  |||| ||||| |||
Db   34991 ATACGAGCCATGTATCATTTCAGAGAGACAGCTACTTCAGAGAGAGCTTGAGAAAACAAGCAGAAGAGCTTGA 34932

Qy    2263 GCGAAGCTCACAGAACAGCAAGCTGCAGAAGCTGCAGGTGGTCGAAGTCAAGAATACAGATAAGTTAC 2322
           || ||||| ||| || || ||  |  ||| || |   |  | ||||| |  || ||||| |||  ||
Db   34931 GTGAAGCTCACAGAAACAACAGTAAGTTGCAGAGTTGCAGAGACTTTAGCATACATAGCTAC 34872

Qy    2323 TCGCCAAGTGCCTCAGGGCTCATCAAGTTCTGAAGGACAGCAAGATCAGCTGCTCCTCTTCCA 2382
           ||||| ||||||||||| |||||||||  | ||||| ||||| || || || ||||| ||||
Db   34011 ACGACTTCTGGGAACGAGCTACCCCTGTTATGAGACTTTAGCATACATAGCTGGGGGCTACTGG 35952
```

```
XX
FH   Key             Location/Qualifiers
FT   CDS             437..2383
FT                   /tag= a
FT                   /transl_except= (pos:1631..1633, Xaa)
FT                   /note= "Xaa = stop codon"
XX
PN   WO20012010-A2.
XX
PD   22-MAR-2001.
XX
PF   14-SEP-2000; 2000WO-EP008994.
XX
PR   16-SEP-1999;  99GB-00021964.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Steimer A, Mittelsten Scheid O,  Paszkowski J;
XX
DR   WPI; 2001-244815/25.
DR   P-PSDB; AAB20321.
XX
CC   Novel transcriptionally silenced genes obtained from Arabidopsis useful
CC   for selecting plants which compared to wild type plant are impaired in
PT   transcriptional gene silencing.
XX
PS   Claim 1; Page 21-22; 36pp; English.
XX
CC   This present sequence is that of a 5' extension of an Arabidopsis
CC   thaliana transcriptionally silent information (TSI) sequence, i.e. TSI-A.
CC   The TSI-A clone was identified by differential mRNA screening. Comparison
CC   of transcription gene expression between an Arabidopsis line carrying a
CC   silent transgene present in multiple copies and its mutant derivative
CC   mom1 impaired in silencing of the transgene revealed 2 cDNA clones which
CC   were expressed in the mutant plants, but not in the parental or wild-type
CC   plants. Both clones were derived from the same family of transcripts
CC   referred to as TSI. Genomic templates encoding TSI are repetitive
CC   elements with mainly pericentromeric location and conserved organisation
CC   among various ecotypes.Transcriptional silencing of genomic TSI
CC   templates is specifically released in the mutant. Transcription of TSI
CC   can be used as a marker to identify a defective silencing pathway in a
CC   plant. The present 5' extension of TSI-A was obtained by PCR. The
CC   sequence includes a putative open reading frame for a 648 amino acid
CC   protein (see Aab20321), but includes an in-frame stop codon. It shows 97%
CC   identity to another isolated 5' extension sequence (see AAF30451). The
CC   invention provides a nucleic acid encoding genetic information which is
CC   silenced in wild-type plants but whose expression is turned on in plants
CC   defective in transcriptional gene silencing, and has the formula: RA-RB-
CC   RC, where RA, RB, RC constitute component nucleotide sequences selected
CC   from the present sequence and chosen given in AAF30450-59. Also provided
CC   are methods of selecting a plant that is impaired in transcriptional gene
CC   silencing, and of producing DNA representing at least part of a gene
CC   necessary to maintain silencing of another gene in a cell or plant
XX
SQ   Sequence 2512 BP; 640 A; 616 C; 634 G; 622 T; 0 U; 0 Other;

Query Match             39.4%;  Score 1230.8;  DB 4;  Length 2512;
Best Local Similarity   74.9%;  Pred. No. 0;
Matches 1644;  Conservative    0;  Mismatches  467;  Indels  83;  Gaps   5;
```

```
Qy    788 ACTCAACCGGTTTACTCGACCGGTTACGCGTTGGTTTAGATTGTGTTTTGGTTTGCTATTA 847
          ||||||   ||   |||||  ||   ||    |||| || |  ||||| |  || |||||
Db    328 ACTCGACCGTTCACTCGATCGCGTACTTGACT--GGTTTAGTGTGTGTTTTATTTGAA 385

Qy    848 CTAACATATTAACGTTTATCTTTGAGTTTC-GTCTGTTTTAGGTTTCATCATGAGTAAC 906
          |||||| |||   |||||| | ||| |||  |||  |||||  ||||| ||||||
Db    386 CTAACATATTGATATTTGGTTTGAGTTACATTCTTTTTCAGGGAATCAATATGAGCAAC 445

Qy    907 TACAGTGGAAAATCCTCATGGACCCTGATTATAATGTGGATGAAGCTAGTCCTGGTCC 966
          ||||||||| |||||||||||||   ||||| |||  ||   || ||| |||| ||||
Db    446 TACAGTGGCCAATCCTCCATGATCGGATTACAACGTCGGAATCCTGAATCCTTGGTCA 505

Qy    967 ACTAGACCGGAGTGAGAGCAACATGTTTACGAGAGCTATAGGAGGATGAATTTGAACGCTCT 1026
```

RESULT 12
AAF30450
ID   AAF30450   standard; cDNA; 2512 BP.
XX
AC   AAF30450;
XX
DT   29-MAY-2001 (first entry)
XX
DE   Transcriptionally silent information clone TSI-A 5' extension.
XX
KW   TSI-A; transcriptionally silent information; gene silencing; ss.
OS   Arabidopsis thaliana.

```
Db   506 |||||| ||||| |||||| |||||| |||||| 565
         ACTAGACCAGAAGAGAGAGCAACACAGCTTATGAGAGCTTCAGAGCCGAGACCCAACGCTCA

Qy  1027 GCCAGCTTCGACGTAATCAACTGAGAGCTGAAATCGCTAGAGGAAAGAGGGCGATGTCCAGT 1086
         ||||||||| |||||||||||||||||  ||  |||||||||| |||
Db   566 GTAGCTCGACGCAATGAAAGGAGGAGCTGAGATTGCTAGAGAGGAAAGAGGAGCCGGATGACCAGC 625

Qy  1087 AGATATGAGCTGATTGATGAGGATATCAAAACTGAGTATGAGCCAGAGTCATGGCCGAAG 1146
         |||||||||  ||||||| |||||||| |||  |||||||||||||||||||| |||
Db   626 AGATATGAGTTGATCGACGAAGAGATATTGACGTCGAGTATGAGCCTGAGTCATGGCCACAGA 685

Qy  1147 GAGACGAAGCTACTGAACAAATCCGACGAGGTTACAGTGGAGGAGTATATCAGATTCTTT 1206
         || ||  |||||||||||||  |||||| |||||| ||||||||||| ||||||
Db   686 GAAACAAAACTGTTGAACAAGCCTGATGAAGAAGAGTTACACGTGGAAGAGTACATCAGACTTTTC 745

Qy  1207 GAGATGAATGACTTCTGGGAACGAGCGTATCCTGATATGAGACTTTAGCCCAGTTGGGG 1266
         |||  |||||||||||||||| | || ||||| |||||||||||  |||||||| | ||
Db   746 GAGCTGAACGACTTCTGGGAGGCGGAGTACCCCTGTTATGAGACTCTAGCCCAGCTTAGG 805

Qy  1267 TTACTGGAGGACGTGCAGCATCTGTTCGAGAGTGTCAT--------- 1305
         |||||||||||| | |||||  |  |||||||| ||
Db   806 CTACTGGAGGACGTACAGCACTTATTCGAGAGTGCCATCTTGAGACGCTGATGTCTTAC 865

Qy  1306 ----------CTGATAAGGAGGACAATGTACCAGTTTCTTTCCACACTGCAAGTGGAAATG 1355
                   ||||  || | ||||  ||| || |||||||||||| ||||||||| ||
Db   866 CCGTACGTCGCTTACAAGAGAACGAAACAATAGAGTTTCTCTCCACTCTGCAAGTGGAGTTG 925

Qy  1356 TATGAGGGACTCACAGACTTTGAGCTGGATACCATGGGGTTAGGCTTCTTGACGTTCTTA 1415
         |||  ||||| ||  | ||| |||  |||| | |||||| || |||||||||| ||||| |
Db   926 TATCAGGGACTTACTGCAGTGCTGGACGATGAAACTGGAGAGGGTGGGATTCTTGACTTTTTCA 985

Qy  1416 GTGGATGAACACGGTACCAGATTTAGATCAAGAAATTGGAAAGACTGTTTGGTTTCCCT 1475
         ||| | |||| |||||||| | |||| |||||||| |||||  |||  ||||||||||
Db   986 GTTAACGAGCAGCGTTACCAGCTATCATCAAGAGCTTGGAAGGATTATTTGGTTTCCCC 1045

Qy  1476 AGTGGAAAGGAACCAACCCCAGGTTGACAGGGAAGAGGTTAAGGAGTTTGTGGCTACT 1535
         ||||||||||||| ||| ||| ||| || ||||||||| ||| ||||||||| | || |
Db  1046 AGTGGAAAGGAACTAAACCCAAGTTCGAAAGGGAAGAGGAAGAGTTGAACGATTTGTGGTTAACC 1105

Qy  1536 ATTGGGAACAATCTACCGTCAAACTCGACGCGGTCCAAGAGCAACCAAATCCGAGTCCT 1595
         ||||||||| | |  |||   ||  || |||||  |||||||||||| || |||| |||
Db  1106 ATTGGGAACCGATTTGGCGCTCAAGGTCTCGAAGGATGCTCAAGAGCAACCAGATTCGAAGCCCT 1165

Qy  1596 GTGATTCGCTACTTTCAGAGCGTCCGGTTGCCAATGTTTTTACTCCGAGGAGTCTACAGGC 1655
         ||||| ||||||| |||| | || |||| |||||| |||  ||| ||||||| ||||||||
Db  1166 GTGATCCGCTACTATCAGCGGTACCAGCTTACAGGGAATGTTCTGTACCCGGAGGAATCTACAGGC 1225

Qy  1656 ACCGTGTCTAACAGACATGAAGATGATAGATTCAGCCGCTTATGAGGCCGCTTATCCGCCTT 1715
         |||||||||||||||||||| |||||||||  ||| |||  ||  |  ||| || ||| | |
Db  1226 ACCGTGTCTAACAGACATGAGGATGATAGAGATTGATTCTGCACTCAAGGGTATTCTCCGGAGA 1285

Qy  1716 ACAAAGAAAGAGAATGTCCTGAGAGAGATCTTAACGACTCACCACCAGTAAATGCCTCTGT 1775
         ||||| | || ||| |||||  | ||| |  |||  ||||||||||| | | || | ||||
Db  1286 ACAAAGGGAGAAGAGGTCCTAAAGCGACCTTAATGATACCACCACCCGGTCATGTCTTCTG 1345

Qy  1776 TTGATCCATCTGTGTGGGTACACCGGAGAATGATGGACTTGGATCACTTGCCGCCGATGCTC 1835
         |||||||  | ||||||| |||| ||| | || |||  |||| || |  |||| | |  | |
Db  1346 TTGATCCCATATGTGGGATACAGGAAGTGGGCGGCGCCACACCGGAGAAGAGGTGCGA 1405

Qy  1836 GGAGCCACTATGCGTGGGGTGCCGGTTGTGCCGTTTGTGCGCCAATTCTGAAAGTTGTCCGCTC 1895
         |||||| | | | ||||||||||||||  | ||| | |||||||  |||||| ||||  ||
Db  1406 GGAGCCCTTTGTGGGGTGCCGGTTGGGTGCCGGTGCGCCACGGGAAGAAGAGGTGCGA 1465

Qy  1896 AAGGAAGTAGGGTTAGCACCCGGAATGATGGACTTGGACTTGCGCCGCCGATGTGAGTTC 1955
         |||| | |  | || |  | ||| | ||||||  ||||   | | |||| | |||| ||
Db  1466 ACGTCTCCAGGGTTTGATCGGCGGTTGATCACTTGGCGTCGTTGTGAGTTT 1525

Qy  1956 TCTGAGTTTGACATGGTTGGCGACTTTCACCGGCTACAGGTTCGAGCATTCATCGATTAGA 2015
         ||| ||||||||  ||||||||| || |  |  || |||| |  | ||  || |||| | |
Db  1526 CTCGAGTACGACATGGTTGGCGATTTCTATCGCTACAAATTCGAGCACTCCCTGACCCCA 1585

Qy  2016 ATCGCCAACATTCTTTCCCCTGCATTTACGCTACTAGGATTCTCGAGGCAGGAACATT 2075
         |  ||||||||| |||  ||||||  ||||||||| ||  ||| ||| |  ||||||
```

```
Db  1586 ACAGCCAACATTTTGCTTCCCTGCATCGAGGCCACAACCATACTTTAGGGTGACGAAACATT 1645

Qy  2076 GACTTCAAGCCTCGCCGTTGAAGATCTTTATTTCGAGGCAGTCGCCAACTGAGCGAGATT 2135
         |||||| ||  |||| | ||||| |||||| |||| |||||| || ||| ||| |||||
Db  1646 GACTTCAGACCTGCCGCGTGATTACCCTCACTTTGAGAGCACTCCACCGACTGATGACAAT 1705

Qy  2136 AGTCACACCGGAAGGAGCTACAATAGAAGATGTTGATGAGACATATGATATAGAGGAGCG 2195
         |||||||| || |  ||| || ||  |||| |||| ||||  ||| |||||||| |||
Db  1706 GTCCCTACGACGGAAGCTACAGAGAGGATGATTTTGCTGAGACGGATGTGGATAGGGAGGAG 1765

Qy  2196 GAGTTTGACACGAGCATGTATCATTCAGTGACGACCATATCCTCCAGCGAGGAGAAAGCAAG 2255
         ||||| ||||| |  |||| |  |  ||| ||||| ||| |||| || |||||  ||  |||
Db  1766 GAGTATGATACGAGCATGTATCATTCAGTGACGACGTACCTCCAGCGCGAGGAGCGCAAG 1825

Qy  2256 AGTTTGAGCGAAGCTCACAGAACAACAGCAAGCTGCAGAAGAAGTGGTGCAAGAAAACAGGAT 2315
         |||||||| |||||||||| | ||| || ||| |  |||||||| |||||||| || |||||
Db  1826 AGCTTGAGTGAAGCTCACAGAACAACAGCAAGCTGCAGAGTGGTGCAAGAAAACAAGAT 1885

Qy  2316 AAGTTACTCGCCAAGTGCCTCAGGCTTCTCGAAGGACAAGATCCAGCTGCTCC 2375
         ||||| |||| ||||||||| ||| | |||| || | |   ||| ||| ||
Db  1886 AGGTTACTTATCAAGTGCTTCAAGGCCATCATCACGTTTCTAACGGACAATAAGTTGTTCTTC 1945

Qy  2376 TCTTCCACTACAACTATTCCGGACTAAGTTTCTGCAAGTGGAAATG 2435
         |||| | || || || ||||| || |||  |||  ||||||||||
Db  1946 TCTTCTACCACAGCTATTCCGGAGGAGGCGCCTCGAGGAGCTGGGG 2005

Qy  2436 GACGCGCCCGAGCCTAGAGAGCAAGAATTCTGCAT----------- 2471
         ||||||||||||||  |||||  | ||| ||| ||
Db  2006 GACGCGCCCAGGCCTGGCCAAGTCACCACCAGCCTGAGCCTAGT 2065

Qy  2472 ------------GTCCCTGCGGCGCGATTCATCATTCGAGCCTCCGTGGAATCTAGG 2513
                     ||| |  || |||||||||| |||||||||| ||| | ||||
Db  2066 GACCGAGTAGTACCACCAGTCCCTGCGGCGCGATTCATCATTCGAGGGAGCTCGGG 2125

Qy  2514 AAGAATAGGAGAACGACACTCGATCTAGCAGCAGCAGAGACTTCTGCAGTCT 2573
         ||||| | |||| | ||||  || |||||||||||||| || | || |||||
Db  2126 AGAAAGAGAGGCTGCACTCGACTGCTAGGTCTGGCAGCAGGTAGACTTTCCCAGTCC 2185

Qy  2574 CGTAGTTTACGGCGACCGCGGTGCGCCAATGAAGTGATCATCAGTCGAGTTCCAG 2633
         ||||||| |||||| | |||||| | |  || |||| | |||| ||| ||||||
Db  2186 CGTAGCTTACGGCGGCCGCGGTGCGCCAGCAGAAGAAGAGAGTCGAGTATCATCAG 2245

Qy  2634 AGCAGGTGCTGGCCGCCAGAGCGTGATGAGACTGCTGAGGAGCTTCTACACGTGATACG 2693
         |||  ||||| |||||| || | || ||| || || ||||||| || || |  ||||
Db  2246 AGCGGGTGCTGGCCGCGCGCGGTGCCGCGGGAAGGAGGAGCTCGAGTGGGGGGGGGACA 2305

Qy  2694 GAACATGGCGGTTCGTCTATGGCTTGGGAGCGCATTGACTACCAACTT 2753
         ||||  | | |  ||| || | ||||| |||||||| || ||||||| |
Db  2306 CAACAGGAGATTCTTCGATGCCTGGGAGCAATCAGCAGCTATTGACGACCAACTC 2365

Qy  2754 CGTTCATTATTCGACTGAGTGAGTTTCAGTGATGATTCCTGATTGTTCTGAATCATCTCTTGT 2813
         |||||| || || ||| | |||||  | ||| ||||||| |||| ||| | ||| ||| |||
Db  2366 CGCTCCTTCTTCCACTGAGGTATGCACTCCACTTCTACTCCACCATTGTAATATACCATCTC---- 2421

Qy  2814 GATTTGTTCTTTATTTGTTCAGTGATTGATTTGGG 2873
         |||||| |||| || | ||||||| |||| |||||
Db  2422 ---TTGTTTTTATTTTGTTTTTGTGTTTTGTCCTGGAGTACTCTTCTTCCAAATTT 2478

Qy  2874 ATTCACACAGTGGACTGTGTGATTTAAGTTTGGG 2907
         ||||||||| || ||||||| |||| ||||| |||
Db  2479 GGTCACACAGTGGACTGTGTGATTTAAGTTTGGG 2512
```

RESULT 13
AAF22288/c
ID   AAF22288   standard; DNA; 92584 BP.
XX
AC   AAF22288;
XX
DT   20-MAR-2001 (first entry)
XX
DE   BAC containing repeats from centromeres 1-4 #11.
XX

KW   Centromere; michrosome; vector; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO20005325-A2.
XX
PD   21-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-US007392.
XX
PR   18-MAR-1999; 99US-0125219P.
PR   01-APR-1999; 99US-0127409P.
PR   18-MAY-1999; 99US-0134770P.
PR   13-SEP-1999; 99US-0153584P.
PR   17-SEP-1999; 99US-0154603P.
PR   16-DEC-1999; 99US-0172493P.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Preuss D, Copenhaver G, Keith K;
XX
DR   WPI; 2000-587529/55.
XX
PT   Recombinant DNA construct comprising a plant centromere, useful for
PT   producing stably inherited michrosomes which can serve as vectors for the
PT   construction of transgenic plant and animal cells.
XX
PS   Claim 102; Page 508-529; 1449pp; English.
XX
CC   The present invention relates to a recombinant DNA construct of a plant
CC   (Arabidopsis thaliana) centromere. The constructs are useful for
CC   producing stably inherited michrosomes which can serve as vectors for the
CC   construction of transgenic plant and animal cells expressing selected
CC   proteins such as hormones, enzymes, interleukins, clotting factors,
CC   cytokines, antibodies, and growth factors
XX
SQ   Sequence 92584 BP; 27840 A; 18113 C; 19835 G; 26796 T; 0 U; 0 Other;

Query Match   38.9%; Score 1215.4; DB 3; Length 92584;
Best Local Similarity   86.8%; Pred. No. 0;
Matches 1405; Conservative   0; Mismatches 161; Indels   53; Gaps 4;

Qy   733 CAACAACTCTCAAAGTAAAAAAAAAGAAAAAAAAGTCAAAACCGACAGTTCACTCA 792
             ||| ||| ||||||||||||||| |||| |||||| |||||||| ||| |
Db   27435 CATCAAAACCGACGAGGAAGAAGTATATAAAAAAGGGATTTGGTCCAAACACTCGACCGGT 27376

Qy   793 ACCGGTTACTCGACCGGTACGCTCGTCTGTTTAGATTGTGTTTTTGGTTTTGCTATTACTAAC 852
             ||||||| ||||||| ||||| |||||||| ||||||||||||||
Db   27375 TTGGTTATACTCGACCTTTTTACTTCGATTGGTTGTGTTTGTGTTTTGGTTTTGCTATT----- 27322

Qy   853 ATATTAACGTTATCTTGAGTTTCGTCGTCTGTTTTGGTTTCATCATGAGTAACTACAGT 912
                   ||||| ||||||||||||| |||||| |||||||||||||||||||||||
Db   27321 ------ACGTTTATCTTTGAGTTTCTGTGTTTTCAGGTTTTCATCATGAGTAACTACAGT 27268

Qy   913 GGAAAATCCCTATGGACCCTGAGTTATATAATCGGATGAAGCTAAGTCCTGGTCCACTAGA 972
             ||| |||||||||||||||||||||| ||||||||||||| ||||| || |||| ||||
Db   27267 GGAGAATCCCTATGGACCCTGAGTTATAATCGGATGAAGCTGAGTCCGGTTAGA 27208

Qy   973 CCGGAGTGAGAGCAACATGTTACGAGAGCTATAGCGCTCTGCAGCT 1032
             ||| || ||||||||||||||||||||||||||||||||||||
Db   27207 CCG----AGAGGCAACATGTTACGAGAGCTATAGCGCTCTGCAGCT 27152

Qy   1033 CGAGGTAATCAATCAGAGAGAGAGCTGAAATCGTAGGAGAAAGAGGGCGATGTCGAGTAGATAT 1092
             ||| ||||||||| ||||||||||| || ||||| ||||||||||| |||| |||||||||||
Db   27151 CGACGTAATCAACATGAGCTACGTGAAATCGCTGAGGGATATCGCTACGTCGAGTAGATAT 27092

Qy   1093 GAGCTGATTGATGAGGATATCAGAGCTACGGAGAGCTATGAGCGCCAGAGTCGAGTCCAAGGACG 1152
             ||||||||||||||||| ||||||||| |||||| ||||||||||||||||||||||
Db   27091 GAGCTGATTGATGAGGATATCTAAACTGACTATGAGCGCCAGAGTCGAGGCGCAAGGACG 27032

Qy   1153 AAGCTGCTACTGAACAAATCCGACGAGGTTACAGTGGAGAGTATATCAGATTCTTTGAGATG 1212
             |||||||||| |||| ||||||||||||||| |||||||||| ||| |||||||||||||||
Db   27031 AAGCTGCTGAACAAACCCGACGAGGTTACAATATCGGAGGAATATATCGACTCTTTGAGATG 26972

Qy   1213 AATGACTTCTGGGAACGAGGTATCCCGATGAGACTTTAGCCCAGTTGGGGTTACTG 1272
             ||||||||||||||| |||||||| ||||||||| || |||| ||||||||||
Db   26971 AATGACTTCTGGGAATGAGGTATCCGTGCTATGAGACTTTAACCAACTGGGGTTACTG 26912

Qy   1273 GAGGACGTGCAGCATCGTTCGAGAAGTGTCATCTG--- 1308
             ||||||||||||||||||||||||||||||||||
Db   26911 GAGGACGTGCAGCATCGTTCGAGAAGTGTCATCTGGAGACACTCATGTCTTACCTTAC 26852

Qy   1309 ------ATAAGGAGGAGACAATGCAGTTTCTTTCCACACTGCAAGTGAAATGTATGAG 1361
                   |||||||||||||||||||||||||||||||| ||||||||| |||||||||
Db   26851 GCTGCCTATAAGGAGGAGACAATGCAGTTTCTTTCCACGCTGCAAGTGGACATGTATGAG 26792

Qy   1362 GGACTCACAGACTTTGAGCTGGATACCATGGGTTAGGCTTCTTGACGTTCTTAGTGGAT 1421
             ||||||||||||||||||||||| |||||||||||||||||||||||||||| |||||||
Db   26791 GGACTCCACAGACTTTGAGCTGGAAACCATGGGTTAGGCTTCTTGACGTTCTCAGTGGAT 26732

Qy   1422 GAACAGCGGTACCAGATTTAGATCAAGAAATTGGAAGAACTGTTTGGTTTCCCTAGTGGA 1481
             ||||||||||| ||| ||| ||||||||| ||| ||||| ||||||||||||||||||
Db   26731 GAACAGCGGTATCAGCTTTCGATCAAGAAGTGAAGAAGTTGTTTGGTTTCCCTAGTGAA 26672

Qy   1482 AAGGGAACCAACCCAGGTTTGACAGAGGAAGAGCTTAAGGATTTGTTGGGTACTATTGGG 1541
             |||||||||||| |||| ||||||||||| ||||||||||||||||||| |||||| |||
Db   26671 AAGGGAACCAAACCCAGATTTGACAGAGGAGGAGCTTAAGGATTTGTTGGGCCACTATAGGG 26612

Qy   1542 AACAATCTACCGTCTAAACTCGACGCGGTCCAAGAGCAACCAAATCCGAGTCCTGTGATT 1601
             ||||||||| |||||||||||| ||||||||||||||||| |||||| ||||| ||| |
Db   26611 AACAATCTATCGTCTAAACTCGGCGCGGTCCAAGAGCAACAAAATCCAGAGTCCTATGATC 26552

Qy   1602 CGCTACTTTCAGCGCTCGGTTGCCAATGTTTTTTACTCCAGGGAGTCTACAGGCACCGTG 1661
             |||||||| ||||| || |||||||| || |||||| | ||||||||||||||||||||
Db   26551 CGCTACTTTGAGCGATCAGTTGCCAACGTCTTTTACTTCCTGGGAGTCTACAGGCACCGTG 26492

Qy   1662 TCTAACAGACATGAAGATGATAGATATAGATCAGCGCTTATAGGATTCTCCGCCTTACAAAA 1721
             |||||| ||||| ||||||||||| ||| || |||| |||| ||||||||| |||||
Db   26491 TCTAACACGGACATGGAGATGATAGATAGATCCGCGCTCACAGGGATTCTCCGCCGTACAAAG 26432

Qy   1722 GGAAAGAATGTCCTGAGAGAGAGATCTTAACGACTCACCACCACCAGTAATGCCTCTGTGATC 1781
             ||||||||||| ||||||||| |||| |||| ||| |||||||||||||| || |||| |||
Db   26431 GGAAAGAATGTCTTGAGAGAGTGATCCATCTCAACACACCACCACCAGTAATGAATGCCTCTTATTGATC 26372

Qy   1782 CATCTGTGTGGGTACATGAAGTGGGCTGGATATAGGAGAGTGAGATGTGAGAGAGCA 1841
             || |||| ||| || |||| |||||||| | || ||| ||| |||||||||||||||
Db   26371 CACCTGTTGTGGATATAGGAGAGTGGGCTGACAAACATTAAGAAGAATGTGAGAGAGCA 26312

Qy   1842 CTATGCGTGGCGTTGGCGTTGTGACGCCAATTCTGAAAGTTTGTTGGAGTTCCGCTCAAGGAA 1901
             |||| ||||||| ||| |||||||||||| | || || |||| |||||| |||||||||||
Db   26311 CTATTTGTGGGTGGCGTTTGACGCGGAACTTGTGGAGTACCGCTCAAGGAA 26252

Qy   1902 GTAGGGTTAGCACCGAGAATGATGACCGATCACTTGGATCATGAGTGCGCCGATGTGAGTTCTTGTCTG 1961
             | |||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||||||||||
Db   26251 CCAGGGTTAGCACCGAGAATGATGACTTGGATCATTGCGCCGATGTGAGTTCTTGTCTGGAG 26192

Qy   1962 TTTGACATGGTTGGCGACTTCACCGTCGAGCATTCATCGATTAGAATCGCC 2021
             |||||||||| ||||||||||||||||||||||| ||||||||||
Db   26191 TTTGACATGGCTGGCGACTTCCACCGTCGAGCATTC------TAGAATCGCC 26138

Qy   2022 AACATTCTTTCCCCTGCTTCATTTTCATTTCGAGGGCAGGAACATTGACTTC 2081
             ||||||||||| || ||||||| |||| ||||||||| || |||||||||
Db   26137 AACATTCTTTCTCCCGCTTCATTCATTTATTTTGAGGGCCAACTGAGGAGATTGACTTC 26078

Qy   2082 AAGCCTCGCGCTTCAGAGTCTTTATTTCGAGGGCGCCAACTGAGGAGATTAGTCAC 2141
             |||||||||||| |||||||| |||||||||||||||||||||||||||| ||||
Db   26077 AAGCCTCGCGCTAGAGAGCTAGCTACAATGATCGCCGAGTCATGGCGCAAGGCGGT 26018

Qy   2142 ACCGAAGGAGCTACAACAGCTACTAGAGCTACACATGATGATAGAGATAT 2201
             ||| || ||||||||||||||| |||| |||||||||||| |||||||
Db   26017 ACCAAAGAGCTACAACAGCTACACAGATGATATAAATGAGGCCGAGTTT 25958

Qy   2202 GACACGAGCATGTATCATTTCAGTGAGCATATATACCTCAGCGGAAAAGCAAGAGTTTG 2261
             |||||||||||||||||||||||||||||||||||||| || |||||||||||||||
Db   25957 GACACGAGCATGTATCATTTCAGTGAGCATATATACCTCCATCGCGGGAAAGCAAGAGTTTG 25898

```
Qy    2262 AGCGAAGCTCACAGGAACAACAGCAAGCTGCAGAAGTGGTGCAAGAAACAGGATAAGTT 2320
           ||||||||||| || |  | |||||||||||||||||||||| |||||||| |||||||
Db   25897 AGCGAAGCTCATAGAAACACACAGCAAGCTGCAGAAGTGGTACAAGAAACATGATAAGTT 25839
```

RESULT 14

AAF22290/c

ID   AAF22290 standard; DNA; 96988 BP.
XX
AC . AAF22290;
XX
DT   20-MAR-2001 (first entry)
XX
DE   BAC containing repeats from centromeres 1-4 #13.
XX
KW   Centromere; michrosome; vector; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO20055325-A2.
XX
PD   21-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-US007392.
XX
PR   18-MAR-1999;   99US-0125219P.
PR   01-APR-1999;   99US-0127409P.
PR   18-MAY-1999;   99US-0134770P.
PR   13-SEP-1999;   99US-0153584P.
PR   17-SEP-1999;   99US-0154603P.
PR   16-DEC-1999;   99US-0172493P.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Preuss D, Copenhaver G, Keith K;
XX
DR   WPI; 2000-587529/55.
XX
PT   Recombinant DNA construct comprising a plant centromere, useful for
PT   producing stably inherited michrosomes which can serve as vectors for the
PT   construction of transgenic plant and animal cells.
XX
PS   Claim 102; Page 550-572; 1449pp; English.
XX
CC   The present invention relates to a recombinant DNA construct of a plant
CC   (Arabidopsis thaliana) centromere. The constructs are useful for
CC   producing stably inherited michrosomes which can serve as vectors for the
CC   construction of transgenic plant and animal cells expressing selected
CC   proteins such as hormones, enzymes, interleukins, clotting factors,
CC   cytokines, antibodies, and growth factors
XX
SQ   Sequence 96988 BP; 29654 A; 18289 C; 18523 G; 30522 T; 0 U; 0 Other;

Query Match   37.6%;   Score 1174.8; DB 3;   Length 96988;
Best Local Similarity 70.8%;   Pred. No. 6.3e-304;
Matches 1391; Conservative   0; Mismatches 237; Indels 584; Gaps   9;

```
Qy     612 ---GTCAAAGAGAGAAGAGAGTGAAGCTGAACCACGGCCACTCGACCGGCGTTCCCTCTA 667
              ||||||||||  ||||| |||||||||| ||  ||||||||||||||||  ||| ||||
Db   67924 CCGAGTCAAAGAAGAAGAGGAGAGTGAAGCTCCATCAACGCCACTCGACCGTTTCCTTCTA 67865

Qy     668 CACATTCAACACTCGACCACGGTGCTACCATCTCCACACCCGCTCTTGTTCACCATACAC 727
           |||||||||
Db   67864 CACATTCAC----------------------------------------------------- 67855

Qy     728 TCGACCAACAACTCTCAAAGTAAAAAAAAAGAAAAAAAAAGTCAAAACCGACAGTTC 787
Db   67854 ----------------------------------------------------------- 67855

Qy     788 ACTCAACCGTTTACTCGACCGCTGGTTTAGATTGTGTTTTGTTTGCTATTA 847
                             |||||||||| |||||| |||||| ||| ||||
Db   67854 ----ACTCGACCGTACAGTCAGTGGTTTTGGTTTGCTTTGTTTGCTGTGTA 67808

Qy     848 CTAACATATTAACGTTTATCTTTGAGTTTCGTCTGTTTTTAGGTTTCATCATGAGTAACT 907
           ||||  ||| |||||||||||||||||||| ||||||| |||||||||||||||||||||
Db   67807 CTAACGCATTAACGTTTATCTTTGAGTTTCATCTGTTTTCAGGTTTCATCATGAGTAACT 67748

Qy     908 ACAGTGGAAAATCCTCATGGACCCTGATTATAATGTGGATGAAGCTAAGTCCTGGTCCA 967
           ||||||||  ||| ||||||||||||||||||||||||||||||||| |||||| |||
Db   67747 ACAGTGGAGAATCTTCATGGACCCTGATTATAATGTGGATGAAGCTGAGTCCTGTTCCG 67688

Qy     968 CTAGACCCGAGTGAGAGCAACATGTTTACGAGCTATAGGATGAATTGAACGCTCTG 1027
           ||||||| |||   ||||||||||||||||||| | ||||||||||||||||| ||
Db   67687 CTAGACCGAGG---GAGCAACATGTTTACGAGCTTTAGGATGAATTGAACGCTCCG 67631

Qy    1028 CAGCTCGACGTAATCAAAGAACGAGCTAGAAATCGCTAGAGGGAAAGAGGGCGATGTCGAGTA 1087
           |||||||||||||||| ||| ||| |||||||||| || | |||||||||||||||||||||
Db   67630 CAGCTCGACGTAATCAGAGAAGCTGAAATCGCCAGAGAAAGAGGGCGATGTCGAGTA 67571

Qy    1088 GATATGAGCTGATTGATGAGGATATCAAAACTGAGTATGAGCCAGAGTCATGGCGCAAGG 1147
           |||||||||| ||||| | |||||| | |||||||||||||||||||||||||||||||
Db   67570 GATATGAGCTACTGATAAGATATCGAAACTGAGTATGAGCCAGAGTCATGGCGCAAGG 67511

Qy    1148 AGACGAAGCTACTGAACAAATCGACGAGGTTACAGTGGAGGAGTATATCAGATTCTTTG 1207
           | |||| |||| |||||||| |||| |||||||||||||||||||||||||||||||||
Db   67510 AAACGAGGCTGTTGAACAAACCGATGAGGTTACAGTGGAGGAGTATATCAGATTCTTTG 67451

Qy    1208 AGATGAATGACTTCTGGGACGAGCTGTATCCCTGATATGAGACTTTAGCCCAGTTGGGGT 1267
           ||||||||||||||| ||||||||||||||     |||||||||||||||||| |||||
Db   67450 AGATGAATGACTTCTAGGGACGAGCTGTATCTTTATTATGAGACTTTAGCCCAGCTCGGGT 67391

Qy    1268 TACTGGAGGACGTGCAGCATCGTTCGGAGAGTGTCATCTG-------------------- 1308
           ||||||||||||||||||||||||||||||||||||||||
Db   67390 TACTGGAGGACGTGCAGCATCGTTCGGAGAGTGTCATCTGGGAGACACTCATGTCTTACC 67331

Qy    1309 ----------ATAGGAGGACAATCGAGTTTCTTTCCACACTCGAAGTGGAAAGT 1356
                     ||||||||||||||||||||||||||||| || | ||||||| |||
Db   67330 CTTACCGCTGTCTATAGGAGGACAATCGAGTTTCTTTCCACGCGTCGCAAGTGGAGATGT 67271

Qy    1357 ATGAGGGACTCACAGACTTTGAGCTGGATACCATGGGGTTAGGCTTCTTGACGTTCTTAG 1416
           |||||||||||||||||||||||||||||||||||||||||| ||||||| |||||  ||
Db   67270 ATGAGGGACTCACAGACTTTGAGCTGGATACCATGGGGTTAGGGTTCTTGGCGTTCGCAG 67211

Qy    1417 TGGATGAACAGCGGTACCAGATTAGATCAAGAAATTGGAAGAACTGTTTGGTTTCCCTA 1476
           |||||||||||||||| ||||| |||||||||| | |||||||||||||||||||||||
Db   67210 TGGATGAACAGCGGTATCAGATTTGATCAAGAAGAAGTTGGAAGAACTGTTTGGTTTCCCTA 67151

Qy    1477 GTGGAAAGGGAACCAACCCCAGGTTGACAGGGAAGAGCTTAAGGATTGTGGGCTACTA 1536
           |||||||| ||||| ||| ||| || ||||||    |||||||| |||||||
Db   67150 GTGGAAAGAGAGAACAACCAAACCCAGATTTGATAGA-------------------- 67116

Qy    1537 TTGGGAACAATCTACCGCTAAACTCGACGCGGTCCAAGAGCAACCAAATCCGGAGTCCTG 1596
Db   67116 ------------------------------------------------------------ 67116

Qy    1597 TGATTCGCTACTTTCAGCGCTCGGTTGCCAATGTTTTTACTCAGGGAGTTACAGGCA 1656
Db   67116 ------------------------------------------------------------ 67116

Qy    1657 CCGTGTCTAACACAGACATGAAGATGAATGATAGCCGCTTATAGGGATTCTCCGCTTA 1716
```

```
Qy     373 ACTTATCACCTTCAAGCTCTTCAGATCTCTCATCTTCATCGTTTCCAACACCGCTGCTCTCCATCCCC 432
               || ||||||  ||| |  | |  ||| ||  || ||||| |||||
Db   68164 ACCTTACCCACTCTATCTCCAAAGATAACACTCGACCTCGCCGCTTCTCACCGGCCTCTCC 68105

Qy     433 CACGGAAAGTCTGTCATCACCTCTCACCTCTCACTCAGTTGACAGGGAAGCAACCAAACT 492
           || || ||||  |||  |  |  | ||| || ||  ||
Db   68104 ATCGTCACCGCCTGCTACTCTCCTCACTCTCCAAAGGAACAACTCCAGCTCCATTCACTCACT 68045

Qy     493 CGACCTCGTCTCTTGCCACTCATAGTCACTCGACTCTTCTCCACCATCTTCATCATCT 552
           ||   |||   |  ||| |  |||   ||  | | ||
Db   68044 TGACCTGGTCTCCCCGCCTCTCCACAATCACTCGATCTCTTCCACCATCTTCATCACCT 67985

Qy     553 CCCTTACTGACCACCGTGGCGTCTCGCGTCGCTCGTCTTGCTCCACATTCACTCATT 611
           |||  ||||||||||  |  ||  | | | | || |  | ||||
Db   67984 ACCATACTGACCACCACCGTCGACCACCGCCGCCATTAAAATCTTCACTCGATTG 67925
```

```
Db    67116  ------------------------------------------------------------  67117
Qy    1717   CAAAAGGAAAGAGATGTCTCGAGAGGATCTTAACGACTCACCACCAGTAATGCCTCTGT  1776
Db    67116  ------------------------------------------------------------  67117
Qy    1777   TGATCCATCTGTGTGGGTACATGAAGTGGGCGCTGACAAACGGCAAGAAGAAGGTAAGAG  1836
Db    67116  ------------------------------------------------------------  67117
Qy    1837   GAGCACTATGCCTGTGGTGGCGTTGTGACGCCAATTCTGAAAGTTGTTGGAGTTCCGCTCA  1896
                                                   ||||||||
Db    67104  ----------------GGAGTACCGCTCA  67104
Qy    1897   AGGAAGTAGGGTTAGCACCGAGAGATGATGACTTGGATCACTTGCGCCGCATGTGAGTTCT  1956
                        |||||||||||||||||||||||||||||||||||||||
Db    67103  AGGAACCAGGGTTAGCACCGAGAGATGATGACTTGGATCATTTGCGCCGCATGTGAATTCT  67044
Qy    1957   CTGAGTTTGACATGGTTGGCGACTTTCACCGCTACAGGTTCGAGCATTCATCGATTAGAA  2016
                                                    |||||||
Db    67043  ----------------------GGTTCGAGCATTCATCGTTTAGAA  67007
Qy    2017   TCGCCCAACATTCTTTTCCCCTGCATTTACGCTACTAGGATTCTCGAGGCAGGAACATTG  2076
                                                    |||||||||||
Db    67006  TCGCCCAACATTCTTCTCCCCTGCATTTACGCTACTAGGATTCTCGAGGAAGGAACATTG  66947
Qy    2077   ACTTCAAGCCTGCGCCTTGAAGATCTTTATTTCGAGGGCAGTCCGCCAACTGAGGAGATTA  2136
                                                 |||||||||||
Db    66946  ACTTCAAGCTTGCGCCTTGAAGATCTTTATTTCGAGGGCCAGTCCGCCAACTGAGGAGTTA  66887
Qy    2137   GTCACACCGAAGGAGCTACAATAGAAGATGTTGATGAGACATATGATATAGATGAGGCGG  2196
                                                 |||||||||
Db    66886  GTCACACCGAGGGAGCTACAATAGAAGATGTTGATGAGACAGAGATGATAGATGAGGCGG  66827
Qy    2197   AGTTGACACGAGCATGTATCATTCAGTCAGTGAGCATATACCTCCAGCGAGAAAGCAAGA  2256
                                                 |||||||||||
Db    66826  AGTTTGACACGAGCATGTATCATTCAGTCAGTGAGCAGCATATACCTCCATCGAGGGAAGCAAGA  66767
Qy    2257   GTTTGAGGGAAGCTCACAGGAACAACAGCAAGCTGAGAAGTGGTCAAGAAAACAGGATA  2316
                                                 ||||||
Db    66766  GCTTGAGTGAAGCTCATAGGAACAACAATAGTAAACTGCAGAAGTGGTGCAAAAAGCACGACA  66707
Qy    2317   AGTTACTCGCCAAGTGCCTCAGGGCTATCAAGTTTCTGAAGGACAAGATCAGCTGCTCCT  2376
                                                 ||||||
Db    66706  AGCTGCTCGCCAAGTGTTTCAAGACTATCAAGTTTCTGACAGATAAGCTGAGTTGTTCCT  66647
Qy    2377   CTTCCACTACAACTATTCCGCAATGACAGCTCCCTCCAGGACATGCGCCTTCGAGGAGATG  2436
                                                 ||||||||||||
Db    66646  CCTCCACTACTACAGCTATTCCCCAGGACAGCTCCCTCCAGGACATGCCCTTCGAGGAGATG  66587
Qy    2437   ACGCGCCCGAGCC------------------TAGAGAGCAGAGAGATTCGCATGTCC  2475
                                                 ||||||
Db    66586  ACGCGCGCCGAGCGCAAGTCTGCGGCCAGCCTAGAGCCTGCGGCGAGAGGATTCCGCAATGTCC  66527
Qy    2476   CTGCGGAGGGCATTCATCATTCCAGCCCTCGTGAATCTAGGAGAGAATAGGAGAACGACACTCA  2535
                                                 ||||||||||
Db    66526  CTGCGGAGGGCATTCATCATTCCAGCCCTCGTGAATCTCGTGAGAGCGGTGCTGGCCGCCCACAGAG  66467
Qy    2536   CTCGATCTAGCAGCAGGAGGAACGCGACTTCTGCAGTCTCGTAGTTTACGCGCACGCCGGTG  2595
                                                 ||||||||
Db    66466  CTCAATCTAGCAGCAGGAGGAGCAGGAGCAGGAGCAGACGCACTCCGACGCGCCGGTG  66407
Qy    2596   CTGGCCCCAATAGAAGAAGAGAGAGTCCAGAGTATCCTCAGAGAGCCGGTGCTGGCCGCCCACAGAG  2655
                                                 ||||||
Db    66406  CTGGCCACGCAAATGAAGGAGTTCGAGTATCCTCAGGGACATCCTCCGACGCGCCGGTG  66347
Qy    2656   CTGATGAGATCGAGTACCCCACATGCTGAGCTGATACGGAACATGCCGGTTCGTCTATGG  2715
                                                 ||||||||
Db    66346  CTGATGAGGTCGAGTACCCCGGTGCTGTGCTCATACCGGAACAAGGCCGGTTAGTCTATGG  66287
Qy    2716   CTTGGGGAGCAATCACAGGCAGCCATTGACAAACAACTTCGTTCATTATTCGACTGAGTA  2775
                                                 ||||||
```

```
Db    66286  CCTTGGGGAGCAATCACAGGCAGCCATTGACAAACAACTTCGTTCATTCTTCGACTGAGGTA  66227
Qy    2776   AGCGGCCTCACTTCACCATTATTAGTAGCATCATCTCTTGTGGATTTGTTCTTATTTGTTTC  2835
                                                 |||||||||||
Db    66226  AGCGGCCTCACTTCACTCATTCATTGTATTATATCATCTTCTGTGGAGTTGTTCTTCATTTGTTTC  66167
Qy    2836   AGTTGATTGGATTTGCTCCTGAGTACTCTTCCAAGTTATTCACACAGTGGACTGTGTGA  2895
                                                 |||||||||
Db    66166  TGTGATTGGATTTGTCTCCTGAGTACTTCTTCCAAGTTATTCACACAGTGGACTGTGTGA  66107
Qy    2896   TTTAAGTTTGGGGAGGAGGCTCAGGAAG--TATGTTGCATTGCATTGCATTTTTAAGTCTCGC  2953
                                                 |||||
Db    66106  TTTAAGTTTGGGGAGGAGGCTCAGGAAGTTGTGTTGCATTGGGTATATTTTGAGTCTGC  66047
Qy    2954   ATTCATCTAAGGCATAGAAAAACCAAAAAAA---TTAAAAATTTCAGAAAATGATTTC  3010
                                                 ||||||||||
Db    66046  ATTCATCTCAGGCATAGAAAAACCAAAAAAAATTGAAAATTTCAGAAAATGATTTC  65987
Qy    3011   ACAAAAAAGAGTGTTCATGTAGTTGCATTACATTTAGGATCAAGTCTAGAGTGTTTCAT  3070
                                                 |||||||||
Db    65986  ACAAAAAAGAGTGTTCATGTAGTTGCATTTACAATTAGGATCGAGTCTAGATTGTTTCAT  65927
Qy    3071   TTAGGATTGTTGCATATGCATAGGGGATAATGATGAGATAGCCTTGTGTAAGCA  3122
                                                 |||||
Db    65926  TTAGGATTGTTACATATGCATAGGGGATAATGATGAGATAGTAGTCTTGTGTAAGCA  65875
```

RESULT 15

AAF30451

```
ID    AAF30451 standard; cDNA; 1997 BP.
XX
AC    AAF30451;
XX
DT    29-MAY-2001  (first entry)
XX
DE    Transcriptionally silent information clone TSI-A 5' extension.
XX
KW    TSI-A; transcriptionally silent information; gene silencing; ss.
XX
OS    Arabidopsis thaliana.
XX
PN    WO200120010-A2.
XX
PD    22-MAR-2001.
XX
PF    14-SEP-2000; 2000WO-EP008994.
XX
PR    16-SEP-1999; 99GB-00021964.
XX
PA    (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI    Steimer A, Mittelsten Scheid O, Paszkowski J;
XX
DR    WPI; 2001-244815/25.
XX
PT    Novel transcriptionally silenced genes obtained from Arabidopsis useful
PT    for selecting plants which compared to wild type plant are impaired in
PT    transcriptional gene silencing.
XX
PS    Claim 1; Page 22; 36pp; English.
XX
CC    This present sequence is that of a 5' extension of an Arabidopsis
CC    thaliana transcriptionally silent information (TSI) sequence, i.e. TSI-A.
CC    The TSI-A clone was identified by differential mRNA screening. Comparison
CC    of transcriptional gene expression between an Arabidopsis line carrying a
CC    silent transgene present in multiple copies and its mutant derivative
CC    mom1 impaired in silencing of the transgene revealed 2 cDNA clones which
CC    were expressed in the mutant plants, but not in the parental or wild-type
CC    plants. Both clones were derived from the same family of transcripts
CC    referred to as TSI. Genomic templates encoding TSI are repetitive
CC    elements with mainly pericentromeric location and conserved organisation
CC    among various ecotypes. Transcriptional silencing of genomic TSI
CC    templates is specifically released in the mutant. Transcription of TSI
```

CC can be used as a marker to identify a defective silencing pathway in a
CC plant. The present 5' extension of TSI-A was obtained by PCR. It shows
CC 97% identity to another isolated 5' extension sequence (see AAF30450).
CC The invention provides a nucleic acid encoding genetic information which
CC is silenced in wild-type plants but whose expression is turned on in
CC plants which are defective in transcriptional gene silencing, and has the
CC formula: RA-RB-RC, where RA, RB, RC constitute component nucleotide
CC sequences selected from the present sequence and those given in AAF30450-
CC 59. Also provided are methods of selecting a plant that is impaired in
CC transcriptional gene silencing, and of producing DNA representing at
CC least part of a gene necessary to maintain silencing of another gene in a
CC cell or plant
XX
SQ Sequence 1997 BP; 537 A; 452 C; 541 G; 467 T; 0 U; 0 Other;

```
Query Match    36.1%;   Score 1127.2;  DB 4;   Length 1997;
Best Local Similarity  75.0%;   Pred. No. 5.4e-292;
Matches 1500;  Conservative  0;  Mismatches 418;  Indels 82;  Gaps 4;

Qy    981 AGAGCAACACATGTTTACGAGCTATAGGAGCTATAGGATGAATTGAACGCTCTGCAGCTCGACGTAA 1040
          ||||||||||| |||  ||||||| || || ||| ||  |  |  ||||||||  |||||||||| |
Db      7 AGGTCAACACAGCTTATGAGAGCTTTAGAGCTGAGACCCAACGCTCAGTAGCTCGACGCAA 66

Qy   1041 TCAAAGAAGAGCTGAAATCGTAGAGAGAGAGGGCGATGTCGAGTAGATATGAGCTGAT 1100
          | |||| ||||||||| | ||| ||||||||||| ||||  ||||||||||||||||||
Db     67 TGAAAGGAGAGCTGAGATTGCTAGAGAGAGAAAGAGAAGAGCAATGACCAGCAGATGAGTTGAT 126

Qy   1101 TGATGAGGATATCAAAACTGAGTATGAGCCAGAGTCATGGCCGAAGGAGGAGACGAAGCTACT 1160
          ||| |||||||| |||  || |||||| |||  | |||||| |||| || ||||| || |||
Db    127 CGACGAAGATATTGACGTCGAGTATGAGACCTGAATCATGGACGAGAAAACGAAGCTGTT 186

Qy   1161 GAACAAATCCGACGAGGTTACAGTGAGGAGTATATCAGATTCTTGAGATGAATGACTT 1220
          ||||||  ||||  || ||||||| |||||||| |||| |  | |||||  |||| |||
Db    187 GAACAAGCCCGATGAAGTTACAGTAGAGAGTACATCAGACTTTTCGAGCTGAACGACTT 246

Qy   1221 CTGGGGAACGAGGTATCCCTGATATGAGACTTTAGCCCAGTTACTGGGAGGACGT 1280
          ||  |||||||||| ||| || |||||||||||||||||| | |||||||||||| |
Db    247 CTAGGGAACGAGGTACCCCGTTATGAGACTTTAGCCCAGCTGGGCTACTGGGAGGACGT 306

Qy   1281 GCAGCCATCTGTTCTGAGAAGTGTCATCTG------------A 1309
          |||| ||| |||| ||||||||| ||||||
Db    307 ACATCACTTATTCGAGAAGGTGCCATCTGAGACGCGTGATGTCTTACCCGTACGTCGCTTA 366

Qy   1310 TAAGGAGGACAATCGAGTTCTTCCACACTGAGTTGAAAGTGAAAATGTATGAGGGACTCAC 1369
          ||| |||| |||| |||||| |||| || |  ||||| ||| |||||||| |||||||| ||
Db    367 CAAGAAGAAACAATAGAGTTTCTCTCCACTCGCAAGTGGAGATGTATCAGGGACTTAC 426

Qy   1370 AGACTTTGAGCTGGATACCATGGGGTTAGGCTTCTGACGTTCTTAGTGATGAACAGCG 1429
          | ||| |||||||||   |||||||| ||| |||| || ||  |||| ||| ||| ||
Db    427 TGCAGATGAGCTGGAGAGTGAAGGGTTGGGGTTCTTGACTTTTTCAGTTTTTCAGTGAT 486

Qy   1430 GTACCAGATTAGATCAAGAAATTGGAGAAACTGTTTGGTTCCCTAGTGGAAATGGACTCAC 1489
          | |||||   | ||| ||| | || |  | ||| |||||| ||| ||||||||||| | ||
Db    487 TTTACCAGCTCATCTATCAAGAGCTTCAAGGATTATTTGGTTTTCCAAGTGGAAATGGGAAC 546

Qy   1490 CAACCCCAGGTTTGAGCTGGAAGAGCTTAAGGATTGTGGGCTACTATTGGGAACAATCT 1549
          |||||| ||||| ||| ||||| || || ||||| ||||  | |||||||||||| ||
Db    547 TAAACCCAAGTTCGAGAGGGAAGAGTTGAAAGATTTGTTGGTTAACCATTGGGAACGATTT 606

Qy   1550 ACCGCTAAACTGCGACGCGTCCAGAGCAACCAAATCCGGAGTCCTGTGATTCGCTACTT 1609
          ||| |||||||| |||  | ||||| |||| || | |  |||  ||||||| |||||| |
Db    607 GGCACTCAACTCTCGCAAGGTCTAAGCAGCTCAAGCCCTGTGATCCGCTACTA 666

Qy   1610 TCAGCGCTCGGTTGCCAATGTTTTTACTTCACAGGGAGTCTACAGGCACCGTGTCTAACAC 1669
          ||||||||| || ||  |||||| | || || ||| | |||||||| | |||||||||||||
Db    667 TCAGCGCTCAGTAGCGAAGTGTTCTGTACCCGAGAATCTACAGGGATCAGCATCGTGTCTAACAC 726

Qy   1670 AGACATGAAGATGATAGTAGATTCAGCGCTTATAGGGAAGACGGCCTTACAAAAGGAAAGAA 1729
          |||||||| ||||||   | ||| || || ||  |||| || |  |||||||  ||||| ||
Db    727 AGACATGAGAGATGATGGATTGTTGCCACTCAAGGGCAGTTCCGAGAACAAGGGGAAGAA 786

Qy   1730 TGTCCTGAGGAGATCTTAAGACTCACCAGTAATGCCTCCGTTGATCCATCCTGTG 1789
          ||||||||||||||||| |  |||||||| | |||  ||||||||||||||||||
Db   1790 TGGGTACTGAAGTGGGCGCTGACAAACGGCCAAGAAGAAGGTAAGAGGCACTATGGCT 1849
```

```
Db    787 GGTCCTAAAGGGCGACCTTAATGATACACCACCGGTTATGCTTCTGTTGATCCACCTGTG 846
          |||  ||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Qy   1790 TGGGTACTGAAGTGGGCGCTGACAAACGGCCAAGAAGAAGGTAAGAGGCACTATGGCT 1849
          ||||||||||||||||||| |||  |||||||||||||||||||||
Db    847 TGGATACAGGAGGTGGGCGCGCACCACCAGAGAAGAAGAAGGTGCGAGAGCCCTTGTGT 906

Qy   1850 GGGTGGCGTTGTGACGCCAATTCGAAAGTTGTGGAGTTCCGTCAAGGAAGTAGGGT 1909
          ||||||| ||||||| ||| ||||| |||| |||||| |||  ||| |||||
Db    907 AGGTGGGGTTGTGACACCGATTCGATTGCATGTGGTGTACCTCCACGTCTCCAGGGT 966

Qy   1910 AGCACCCGAGAATGATGACTTGGATCACTTGCGCGATGTGAGTTCTCTGAGTTTGACAT 1969
          |||||||||||| | || |||||||||||||| | || ||||| |||||||||||||||
Db    967 TGATTGGATTTAGATGGATTTAGATCATCACTTGCCTCGTCGTGTGAGTTTCTAGACGACAT 1026

Qy   1970 GGTTGGCGACTTTCACCGCTACAGGTTCGAGCATTCGATTAGAATCGCCAACATTCT 2029
          ||||||||| ||| | |||||| | ||||||||  | || | ||   |||||||| |
Db   1027 GGTTGGCGATTTCTATCGCTACAAATTCGAGCACTCCCTGACCCGGACACAGCCAACATTTT 1086

Qy   2030 TTTCCCCTGCATTTACGCTACTAGGATTCTCGAGGGCGAGGGACATTGACTTCAAGCCTGC 2089
          | ||||||||  | | ||| |  |||||| ||||| || |  ||||||| ||||||||||
Db   1087 GCTTCCCTGCTGCGAGGCCAACCATATCTTCAGGGTGAGAACATTCAGGGTCGACCTGC 1146

Qy   2090 GCTTGAAGATCTTTATTTCGAGGGGCAGTCCCGCCAACTGAGAGATAGTCACACCGAAGG 2149
          || ||| |||||| || || ||  | ||||||| || ||||| ||| || ||||||||
Db   1147 GCGTGATTACCTCTACTTTGAGAGCGCGTCCACCGACTGATGCAATGCCCGACGCGGA 1206

Qy   2150 AGCTACAATAGAAGATGTTGATGAGCACATATGTATATGAGGGCGGAGTTTGACACGAG 2209
          ||  ||||   ||||||| |||  |||| ||||||||| ||||| ||||  ||| ||||
Db   1207 AGTTACAGAGGATGATGATATTGCTGAGACCGGATGATAGGGAGGAGTATGATACGAG 1266

Qy   2210 CATGTATCATTTCAGTGAGCATATACCTCCAGCGCGAGGAAAGCAAGAGTTTGAGCGAAGC 2269
          ||||||||||||||||||||||     || || |  || ||||||  |  |||||| ||||
Db   1267 CATGTATCATTTCAGTGAGCATGAGCAACGTTTCAGGTGAGTGAGCTTGAGTGAAGC 1326

Qy   2270 TCACAGGAACAACAGCAAGCTGCAGAAGTGGTGCAAGAAGGAAGTAAGTACTCGCCAA 2329
          ||||||  ||||||| |||||| |||| |||||  || || | || || ||| |  ||
Db   1327 TCACAGAACACACAGTAAGTTGCAGAGTGGTGTGCAAGAAACAAGATAGGCTACTTATCAA 1386

Qy   2330 GTGCCTCAGGGCTATCAAGTTCTGAAGGACAAGATCAGCTGCTCCTCTTCCACTACAAC 2389
          |||| |||| ||||| | ||| |||| |||||||| ||||| | |||||||||| |||
Db   1387 GTGCTTCAAAGCCATCACGTTTCTGACGGACAAGATAAGCTGCTTCCTCTTCCACCAGC 1446

Qy   2390 TATTCCGCAATGACAGCTCCCCTGAGGACATGACGC------------------ 2440
          ||||||||  | |||| | || |||| |||||||||
Db   1447 TATTCCGCAGGGGAGCATCCTCAAGGAGCATCCTTCAGGAGATATGACGCGCCCAGCGCC 1506

Qy   2441 ---------------GCCCGAGCCTAGAGAGCAGCAGAAGATTCT 2467
                         || |||||| || |||| ||| |||  ||||
Db   1507 AAGTCATCACAGGCCTGAGCCAAGTCACCAGGCCTAGTGCCAGTCAGTAGTCCC 1566

Qy   2468 GCATGTCCCTGCGAGGCATTCATCATTCGAGCCTGCGAGGCCTCGTGAATCTAGGAGAAC 2527
          |||||||||||| |||||| ||| | |||||||| ||| ||||| |||| ||||||||
Db   1567 ACCAGTCCCTGCAAGGCATCATCCCTCAAGGAGCATCCCGGAGCTCGGAGAAGAAGAAGGC 1626

Qy   2528 GACACTCACTCGATCTAGCAGCAGGAGCAGAGAGAGAGGTCAGTATCCTCAGAGCGTGGCCG 2587
          ||||||| || || || | ||||| |||||||  | |||| || |  ||||| ||||  ||
Db   1627 TGCACTCGCTCGGCTCGGCAGGAGCAGAGTACGACAGAGTACGACGCGAGTAGTCCGA 1686

Qy   2588 CCGCGGTGCTGGCCGCAATAGAAGAGAAGAGGTCGAGTATCCTCAGAGCGTGCTGGCCG 2647
          || | |||||||||||  | |||||||||||||||||||| ||||||||||||||||||
Db   1687 CCGGTGTGCTGGCCGCCGCAGCAGAGAAGAAGAGGTCGAGTATCATCAGAGCGTGCTGGCCG 1746

Qy   2648 CCACAGAGCTGATGACGAGATCGAGTACGCTGAGCTCGGACATGCGGGGCCGGTTC 2707
          | || || || ||| || ||| ||||| |  |||  |||||||||||| | |||||
Db   1747 CGACGAAGGAGACGAGGTCGAGTACCCCCACGCGGGGAAGCTGAGACACAACAGGAGATTC 1805

Qy   2708 GTCTATGGCTTGGGAGCAATCACAGGCAGCCACTTCGTTCATTATTCGA 2767
          | | || || ||||||||||||||  |  ||  |  ||||| |||| || |||
Db   1806 TTCGATGCGTGGGAGCAATCACATGCACGACCAAGCTTATTGACGACCATGCCTTCTCCA 1865

Qy   2768 CTGAGGTAAGCGCCTCACTTCACCATTATATTATATTATATACCATTCTTGTGATTTGTTCTTTAT 2827
          |||||||| || ||||| |||| ||| | ||| | |||| |||||| | |||      ||||| |||
Db   1866 CTGAGGTATGCACCCTCACTTCACTTCCACCATTGTAATATACCATCTCTTG------TTTTTTAT 1917
```

```
Qy    2828 TTTGTTTCAGTGATTGATTTGTCCTGAGTACTCTCTTCCAAGTTTATTCACACAGTGGA 2887
           |||||||||  |||||| |||||||||||||||||||||||| |||  |||||||||||
Db    1918 TTTGTTTTTGTGATGTGTTTGTCCTGAGTACTCTCTTCCAAATTTGGTCACACAGTGGA 1977

Qy    2888 CTGTGTGATTTAAGTTTGGG 2907
           |||||||||||||||||||
Db    1978 CTGTGTGATTTAAGTTTGGG 1997
```

Search completed: September 9, 2004, 04:01:20
Job time : 1136 secs

```
                    GenCore version 5.1.6
         Copyright (c) 1993 - 2004  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:     September  9, 2004, 03:36:11 ; Search time 239 Seconds
                                          (without alignments)
                      7249.191 Million cell updates/sec

Title:          US-10-088-384A-27
Perfect score:  3122
Sequence:       1 ttcatatattcgacctcttc.........atgagatagccttgtaagca 3122

Scoring table:  IDENTITY_NUC
                Gapop 10.0, Gapext 1.0

Searched:       682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters:       1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    Issued Patents NA:*
        1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
        2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
        3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
        4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
        5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
        6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	751.6	24.1	1857	4	US-09-322-478-24	Sequence 24, Appl
2	90	2.9	7218	1	US-08-232-463-14	Sequence 14, Appl
3	63.8	2.0	1166	4	US-09-072-596-323	Sequence 323, App
4	63.8	2.0	1166	4	US-09-072-967-328	Sequence 328, App
c 5	63.8	2.0	7218	1	US-08-232-463-14	Sequence 14, Appl
6	55	1.8	53526	3	US-08-658-136-2	Sequence 2, Appli
7	55	1.8	53577	3	US-08-658-136-1	Sequence 1, Appli
c 8	45.4	1.5	6583	4	US-10-204-708-25	Sequence 25, Appl
9	42.6	1.4	318	5	US-09-165-264-12	Sequence 12, Appl
c 10	41.6	1.3	32207	2	US-08-770-379-20	Sequence 20, Appl
c 11	41.6	1.3	32207	4	US-09-230-371A-20	Sequence 20, Appl
c 12	41.6	1.3	32207	4	US-09-621-976-2813	Sequence 2813, Ap
13	40.6	1.3	832	4	US-09-257-179-22	Sequence 22, Appl
14	40.4	1.3	487	4	US-09-165-264-7	Sequence 7, Appli
15	40.2	1.3	320	5	US-09-165-264-11	Sequence 11, Appl
16	40	1.3	320	3	US-09-007-005-17	Sequence 17, Appl
c 17	39.8	1.3	289	3	US-09-244-796-17	Sequence 17, Appl
18	39.8	1.3	289	1	US-09-165-264-14	Sequence 14, Appl
19	39.6	1.3	320	1	US-08-458-568A-11	Sequence 11, Appl
c 20	39.6	1.3	12001	1	US-07-867-106-2	Sequence 2, Appli
21	39.4	1.2	5852	4	US-09-482-273-50	Sequence 50, Appl
22	39	1.2	867	1	US-08-971-395-1	Sequence 1, Appli
23	39	1.2	3182	3	US-08-413-135-1	Sequence 1, Appli
24	39	1.2	3183	4	US-09-621-976-2813	Sequence 2813, Ap
25	38.8	1.2	832	3	US-09-020-956-102	Sequence 102, App
c 26	38.6	1.2	470	3	US-09-030-607-102	Sequence 102, App
c 27	38.6	1.2	470	4	US-09-439-313-102	Sequence 102, App
c 28	38.6	1.2	470	4	US-09-352-61AA-102	Sequence 102, App
c 29	38.6	1.2	470	4	US-09-232-149A-102	Sequence 102, App
c 30	38.6	1.2	470	4	US-09-159-812-102	Sequence 102, App
c 31	38.6	1.2	470	4	US-09-636-215-102	Sequence 102, App
c 32	38.6	1.2	470	4	US-09-685-166A-102	Sequence 102, App
c 33	38.6	1.2	470	4	US-09-115-453-102	Sequence 102, App
c 34	38.6	1.2	470	4	US-09-688-489-102	Sequence 102, App
c 35	38.6	1.2	8961	4	US-10-204-708-80	Sequence 80, Appl
c 36	38.6	1.2	11558	5	PCT-US93-0625-23	Sequence 23, Appl
c 37	38.4	1.2	3001	4	US-09-539-333D-195	Sequence 195, App
c 38	38	1.2	5476	4	US-10-204-708-81	Sequence 81, Appl
c 39	37.8	1.2	319	3	US-09-165-264-8	Sequence 8, Appli
c 40	37.8	1.2	320	3	US-09-165-264-13	Sequence 13, Appl
c 41	37.8	1.2	746	3	US-09-013-810-1	Sequence 1, Appli
42	37.8	1.2	392000	4	US-10-027-983-11	Sequence 11, Appli
43	37.6	1.2	950	3	US-08-991-789A-6	Sequence 6, Appli
44	37.6	1.2	950	4	US-09-062-451-6	Sequence 6, Appli
45						

ALIGNMENTS

```
RESULT 1
US-09-322-478-24
; Sequence 24, Application US/09322478
; Patent No. 6331662
; GENERAL INFORMATION:
;  APPLICANT: Wright, David A.
;  APPLICANT: Voytas, Daniel F.
;  TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
;  FILE REFERENCE: P-1065 ISURF Plant Retroelement
;  CURRENT APPLICATION NUMBER: US/09/322,478
;  CURRENT FILING DATE: 1999-05-28
;  EARLIER APPLICATION NUMBER: 60/087125
;  EARLIER FILING DATE: 1998-05-29
;  NUMBER OF SEQ ID NOS: 41
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
;  LENGTH: 1857
;  TYPE: DNA
;  ORGANISM: Arabidopsis thaliana
US-09-322-478-24

Query Match         24.1%;   Score 751.6;  DB 4;  Length 1857;
Best Local Similarity  65.3%;   Pred. No. 1.9e-210;
Matches 1250;  Conservative 0;  Mismatches 564;  Indels 100;  Gaps 6;

Qy   898  ATGAGTAACTACAGTGGAAAATCCTCTATGGACCCTGATTATAATGTGGATGAAGCTAAG  957
              ||||||| |||||||| ||| | ||| ||| || ||| ||||||| |||||||||
Db     1  ATGAGCAATTACAGTGGCAGTTCTTCTTGTTGATCCTGACTACAACATGGATGAGACAGAA  60

Qy   958  TCCTGTCCACTAGACCGGAGTGAGAGCAACATGTTTACGAGAGTCAACATGGGATGAATTT  1017
              || || || |||||  || | |  ||||||    ||||||  |||| |||||| ||||
Db    61  TCGTCATCTTCAAGGCCAGAGGCCAGAGAGAGAACAGAGAGATACGAAAGTTTCAG-----  110

Qy  1018  GAACGCTCTGCAGCTCGACGTCGACGTAATCAAAGAAGAGAGAGCTGAAATCGCTAGAGGAAAGAGGGCG  1077
              ||||||||||||||| |||||| |||||||| ||||
Db   111  -------------------AAGGAAAGCTGAGATAGCCCGAGGAAAGAGAGAGCG  144

Qy  1078  ATGTCGAGTAGATATGAGCTGATTGATGAGGTATCAAAACTGAGTATGAGCCCAGAGTCA  1137
              |||   ||| ||||||||| ||| |||  ||| | ||||  ||||  |  || |||
Db   145  ATGAGAGAGGATATGAGCTTATAGACGAAGATCTGGACGACGAGTACATGCCTGAACAG  204

Qy  1138  TGGCGCAAGGAGACGAAGCTACTGACAGAACAAATCCACGAGGTTACAGTGGAGGAGTATATC  1197
              ||| ||||||  |||||||||| || ||||  || ||||||  |||||| |||| |||
Db   205  ACTCCAGAGCTACCAAACTTCTGCACAAGCCCGACATATTGCCTGCTGAGGAATATGTT  264

Qy  1198  AGATTCTTTGAGATGAATGAACTACTTCTGGGGAACGAGGTATCCCTGATATGAGACTTTAGCC  1257
              |||||||| |||||| ||||  |||||| |||||||| | |||||  ||||||||||||
Db   265  AGGCTTTTCAAGTCGAATGAGTTCTGTAGCACGAGTATCCTTGCTGACCTCACTTGCA  324

Qy  1258  CAGTTGGGGTTACTGGAGGACGTGCAGCATCGTTCGAGAAGTGTCATCTG-------  1308
```

RESULT 2
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 3047Z/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs

[Sequence alignment data: paired Qy (Query) and Db (Database) nucleotide sequence lines with match bars, numbered by base position, spanning positions 325–2768 (Db) and 325–2366 (Qy).]

```
;    TYPE: nucleic acid
;    STRANDEDNESS: single
;    TOPOLOGY: linear
;    IMMEDIATE SOURCE:
;      CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match          2.9%; Score 90; DB 1; Length 7218;
Best Local Similarity   4.5%; Pred. No. 1.8e-15; Indels   0; Gaps   0;
Matches  18; Conservative  253; Mismatches 133; Indels   0; Gaps   0;

Qy     50 AAGCCGCCATCGTTCCCTCACAACTCTCACTCGACCACCGCCGCTCTCCACTTACT 109
          ||   ||   ||| |::|||::::: ::: :::  ::  :::::::::::::::: ::
Db   1053 AGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1112

Qy    110 CGGCTTCATCGCTCTCATCGCCATCTCTCAACATACTCGACCTCGCGATATCACTCGAGC 169
          ::::::::::::||||||||||| :::::: :  : ::: :::: :::: :: ::: :::
Db   1113 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1172

Qy    170 TCGCGCCGTTCTCACCGGCCTTCCATCGTCACCGCCTGCTCCCTGTCCTCCAAGGAAACAAC 229
          ::  : ::: :::::::::::::::::: :::::::::::::::::::::::::::::::::
Db   1173 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1232

Qy    230 TCGAGCTCTCCATTCACTCGACCTCTACCACCAAGGCCGCTTCACCACTTCTAGC 289
          :::: :::: ::: :::::: ::: ::::: ::::::::::::::::::::::::::
Db   1233 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1292

Qy    290 TCTTAACCACTCGACCACCTTCACCATCAACCAATCAAATGTTTTCTCCCTCCATTAAAG 349
          ::::::::::::::::::::::::::: :::::: :::::::::::::::::::::::::::
Db   1293 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1352

Qy    350 CTTGACATACTCGACCGCTGAACACTTATCACCTTCAAGCTCCTCATCTTCTTCATCGTTT 409
          ::::  :: ::::::::::::::::::: :::::::::::::::::::::::::::::::::
Db   1353 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1412

Qy    410 CCAACACGCTGCTCTCATCCCCACGAAAGCTTGTCATCACCT 453
          :::: :::: :::::::: :::: |||||  |||||||||
Db   1413 YYYYYYYYYYYYYYYGTACCAAATTCTTCTATCTCTT 1456
```

```
RESULT 3
US-09-072-596-323
;  Sequence 323, Application US/09072596
;  Patent No. 6458366
;  GENERAL INFORMATION:
;    APPLICANT: Reed, Steven G.
;    APPLICANT: Skeiky, Yasir A.W.
;    APPLICANT: Dillon, Davin C.
;    APPLICANT: Campos-Neto, Antonia
;    APPLICANT: Houghton, Raymond
;    APPLICANT: Vedvick, Thomas S.
;    APPLICANT: Twardzik, Daniel R.
;    APPLICANT: Lodes, Michael J.
;    APPLICANT: Hendrickson, Ronald C.
;    TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF    TUBERCULOSIS
;    NUMBER OF SEQUENCES: 350
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: SEED and BERRY LLP
;      STREET: 6300 Columbia Center, 701 Fifth Avenue
;      CITY: Seattle
;      STATE: Washington
;      COUNTRY: USA
;      ZIP: 98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/09/072,596
;      FILING DATE: 05-MAY-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
```

```
;      NAME: Maki, David J.
;      REGISTRATION NUMBER: 31,392
;      REFERENCE/DOCKET NUMBER: 210121.417C9
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (206) 622-4900
;      TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 323:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 1166 base pairs
;      TYPE: nucleic acid
;      STRANDEDNESS: single
;      TOPOLOGY: linear
;    MOLECULE TYPE: Genomic DNA
US-09-072-596-323

Query Match          2.0%; Score 63.8; DB 4; Length 1166;
Best Local Similarity  26.7%; Pred. No. 2.8e-08;
Matches 190; Conservative 136; Mismatches 384; Indels   1; Gaps   1;

Qy      2 TCATATATTCGACCTCTTCTTCACTCTTCGCATCCAAAAGACACAAGCACGCCATCG 61
          |:|  || |  ||||: :: :::  :: :::: ::  :: :: :::  ::::::|:|
Db    373 KSAAMTSMKMGGTSYCTMTYCNNGGASTAMTYNMCCCGWAYCKSCWAYCCCTCGTCA 432

Qy     62 CTTTCCCTCACAACTCTCACTCGACCACCGCCGCTCTCCACTTACTCGGCTTCATCGC 121
          :: ||| |:: ::::: ::: :::::: :::: ||| ::: ::||: :: :| :||| |
Db    433 TYCCMCMCSGSGYCCTCAMNCCACCYTGNGYCCCTCCMKMTCYCAYTCMNTCCGGTWCC 492

Qy    122 TCTCATCGCCATCTCTCAACATACTCGACCTCGCGGATATCACTCGAGCTCGCCGCTTCTC 181
          |:|:|::| ||| :: :: ::  : ::::: :||:: :||:||| ::|:: :||::: :|
Db    493 TNTMNNCCCSCNCRYCTCAMCNCTKSGKCACCNATMYCCSACKCHTCTMCYMCSCAKNMT 552

Qy    182 ACCGCCTCTCCATCGTCACCGCCTGCTCCCTCCTCCAAGGAAACAACTCGAGCTCTCCA 241
          :::::|| :: :::|::::: :|::|| ::::::::| :: :: :: ::::||: ::|:
Db    553 TCCCCTCNCCTYTNNCCAMCMCSCTCTMTCMAACTCKKCCCGGYCKCNCMYCTCKCCCAY 612

Qy    242 TTTCACTCACTGACCTCTACCACCAAGCCGGCTTCACCGTTCATCACTC 301
          ||| ::|:|: |::|::  : :| :: |:||:::|::::|  :||  |||
Db    613 NMAACCKKTYCYWCNWYCYMYCKCKCAGWYKNMCTCCWACTCTMYNTTTCTCTNKCCCM 672

Qy    302 GACCACCTTCACCATCAACCAATCAACAAATCAAATCGTTTTCTCCATTAAAGCTTGACATATCTC 361
          :||| ::: :: ::::| :  ::|:  :: ::: |:|: : :: ::: ::: ::|  ::|:|:
Db    673 KACCKNTTCTCWCSCCCCCCACAKAYMCYAWCWTMTCMCTCKCACSCCCYYCNNYCCMNM 732

Qy    362 GACCGCTGAACACTTATCACCTTCAAGCTCCTCATCTTCTTCATCGTTTCCAACACCGCTG 421
          :|::|  |||| ::| :: ::::|||:::: :| ::|::::|||||::::::::: :::::
Db    733 CWCMTCWCTWNAKCANCNTTCTTCTCTCMMYMTMACKCWCNNTCNCCKSGACCYYTCCAC 792

Qy    422 CTCTCATCCCCACGAAAGCTTGTCATCACCTCCACTTCACTCACTCCAGTTCACTCGATTCA 481
          :| ::|||:::| ::|| | |: ::::: ::::::| ::||:|: :: :: :::||::||||
Db    793 TRMKCCKMTCTCCTTMCKCCYMWCNTCCMKYNCCCTCCNMTCMTCKYTCCTCNCNMRYCY 852

Qy    482 GCAACCAAACTCGACCTCGTCTCTTCTGCCACTCATAGTCACTCGATCTCCTCCTCACCA- 540
          | ::| ::| :::|| ::: :::::| ::|:: :::| :||::::::||: ::|::::|:
Db    853 YYAKCAKCNMCTCCCANKMCAKCTKCTCCCCCAKMKSACNCKKCCCWCCCCTATCCWC 912

Qy    541 TCTTCATCATCTCCCTTACTGACCACCGTGCGTCTCGCCTCCACCATTGCCATTTAAAAG 600
          ||  ::::: :::::: :: :||::|| :::: |: ::| ::|:|::||:: :::|:: ::
Db    913 TCTCWCTYATCTCKCTCWCNYCMYMKMCACNCKCYAYTCNACTMNMWNCCANCNCTCTCT 972

Qy    601 CTCACTCGATTGTCAAAGAGAAGAAGAAGAGTGAAGCTCAACCACTGACCACGCCGGTTT 660
          :| ::| :::|| :: ::|| :: :: :: :  :::| :: ::||:| ||:  :|::|::
Db    973 NYCTCWCKACGTYCKCCKCCTMCKCNMCNRWCTYRCCTCKKCCNCCRNCKNMCMKCTMCT 1032

Qy    661 CCCTCTACACATTCAACACTCGACCACGGTGCTACCACTTCCACACCCGCT 711
          :: :::|:::|:  :  :: :|||| ::::  :: :::: ::|::|:::| :
Db   1033 CTCCWMKMTCCCWCCCATCTMMKSTCTWCNCMTCCCTCNCCCYNYNTKCY 1083
```

```
RESULT 4
US-09-072-967-328
;  Sequence 328, Application US/09072967
;  Patent No. 6592877
;  GENERAL INFORMATION:
```

```
; APPLICANT:  Reed, Steven G.
; APPLICANT:  Skeiky, Yasir A.W.
; APPLICANT:  Campos-Neto, Antonio
; APPLICANT:  Dillon, Davin C.
; APPLICANT:  Houghton, Raymond
; APPLICANT:  Vedvick, Thomas S.
; APPLICANT:  Twardzik, Daniel R.
; APPLICANT:  Lodes, Michael J.
; APPLICANT:  Hendrickson, Ronald C.
; TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION:  AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES:  355
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE:  SEED and BERRY LLP
;   STREET:  6300 Columbia Center, 701 Fifth Avenue
;   CITY:  Seattle
;   STATE:  Washington
;   COUNTRY:  USA
;   ZIP:  98104-7092
; COMPUTER READABLE FORM:
;   MEDIUM TYPE:  Floppy disk
;   COMPUTER:  IBM PC compatible
;   OPERATING SYSTEM:  PC-DOS/MS-DOS
;   SOFTWARE:  PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/09/072,967
;   FILING DATE:  05-MAY-1998
;   CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
;   NAME:  Maki, David J.
;   REGISTRATION NUMBER:  31,392
;   REFERENCE/DOCKET NUMBER:  210121.411C9
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  (206) 622-4900
;   TELEFAX:  (206) 682-6031
; INFORMATION FOR SEQ ID NO:  328:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  1166 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  Genomic DNA
US-09-072-967-328
```

```
Query Match        2.0%;   Score 63.8;   DB 4;   Length 1166;
Best Local Similarity  26.7%;   Pred. No. 2.8e-08;
Matches  190;   Conservative  136;   Mismatches  384;   Indels  1;   Gaps  1;

Qy      2 TCATATATTCGACCTCTTCTTCTCATTCTTGCATCCAAAAGACACAACAAGCCGCCATCG 61
            ::| ::|  :|:  |:|:| | |||  :|: |||||  ||  |  ||||:||||
Db    373 KSAAMTSMKMGGTSTYCTMTYCNNGGASTAMTYNMCCCCGWAYCKSCWAYCCCTCGTCA 432

Qy     62 CTTTCCCTCACAACTTCACTGACCACCGCGCCGCTCTCACTTACTCGGCTTCATCGC 121
            |||||  || ||| |||| | || |||  :|: |||  ||||:|||| ||| |||:
Db    433 TYCCMCMCSGSGVCCTCAMNCCACCYTGNGYYCCCTCCMRMTCYCAYTCMNTCCGGTWCC 492

Qy    122 TCTCATCGCCATCTCTCAACATACTCGACCTCGGGATATCACTCGAGCTCGCCGCTTCTC 181
            |  ||:|  ||||||  ::||: |  : :||:  |:|||  ::|:|:| : |:| :|:
Db    493 TNTMMNCCCSCNCRYCTCAMCNCTKSGKCACCNATMYCCSACKCHTCTMCYMCSCAKNMT 552

Qy    182 ACGGCCTCCCATCGTCACGGCCTGCTTCCCCTGTCCCTCTCCAAGGAAACAACTCGAGCTCTCCA 241
            ||  |||  |||||  |||| ::||| |  : |||  ||||  |:|: |  ||  |:|| |:
Db    553 TCCCCTNCCCYTNNCCAMCMCGSCTCMTCMAACTCKCCCGGYCKNCMYCTCTCKCCAY 612

Qy    242 TTTCACTCACTGACCTCACCACCAGCCGGCTTCACCACTTCTAGCTCTTAACCACTC 301
            ||| ||| |||| ||| | :||| ||||||||:| |||| ||:|: |||| | ||||
Db    613 NMAACCKKTYCWCNWYCYMYCKCKCAGWYKNMCTCWACTCMYNTTTCTCTCNKCCCM 672

Qy    302 GACCACCTTCACCATCAACCAATCAACAAATCGTTTCTCTCATTAAAGCTTGACATACTC 361
            ||||  :|:|| || | ||| ||| || ||:|:||  :|||: |||  |  | |||
Db    673 KACCKNTTCTCWCSCCCCCCACAKAYMCYAWCMTMTCCMCTCKACSCCCYYCNNNYCCMNM 732

Qy    362 GACCGCTGAACACTTATCACCTTCAGAGCTCCTCATCTTCATCGTTTCCAACACCGCTG 421
```

```
Db    733 CWCMTCWCTWNAKCANCNTTCTTCTCTCMMYMTMACKCWCNNTCNCCKSGACCYTCTCAC 792
            :|:  |:|  ||  |  ||  ||||||:| |  || |:| ||| :|||  | :|:  |||
Qy    422 CTTCATCCCCACGAAAGCTTGTCATCACCTTCATCACCAGTTCACTCGATTCA 481
            :||: ::|  ||  |||  ||||  |:| |  | :||:
Db    793 TKMKCCKMTCTCCTTMCKCCYMWCNTCCYMWCTCCMKYNCCCTCCNMTCMTCKYTCCTCNCNMRYCY 852
            :|| |:|: ::|||
Qy    482 GCAACCAAACTCGACCTCGTCTTCTTGCCACTCATAGTCACTCGATCTCTCCTCACCA- 540
            |:|:   ||||| | :|||  :|  |  :|:|  ||:|:|  |:| :||:|||||||
Db    853 YYAKCAKCNMCTCCCCANMCAKCTKCTCCCCCAKMKSACNCKCCCWCCTCCTATCCWC 912
            |:|:|  ||||:||||
Qy    541 TCTTCATCATCCCTTACTCGACCACCGTGGTCTCGCTCGCTCCACCATTGCCATTAAAAG 600
            |||: |:|||:|||:|| | | |:  ||| || |  |  :|: |:
Db    913 TCTCWCTYATCTCKCTCWCNCYCMYMKMCACNCKCYAYTCNACTMMWNCCANCNCTCTCT 972
            :|  :|:|
Qy    601 CTCACTCGATTGTCAAAGAGAAGAAGAGTGAAGCTCAACCACCGCCACTCGACCGCGTTT 660
            :|  :|:|:  | :|:| | |:||  || | |  : |:|||  | |  | ||| |||
Db    973 NYCTCWCKACGTYCKCCKCTMCKCNTMCNRWCTYRCCTCKKCCNCCRNCKNMCMKCTMCT 1032
            :|| ||:|||:| |||
Qy    661 CCCTCTACACATTCAACACTCGACCACGGTGCTACCATCTCCACACCCGCT 711
            |||| |:| |:|:|||||||| |||  | :|:|||  ||| | :||||  |||
Db   1033 CTCCWMKMTCCCWCCCATCTMMKSTCTCWCNCMTCCCTCNKCCYNYNTKCY 1083
```

```
RESULT 5
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
;   APPLICANT:  DORNER, F.
;   APPLICANT:  SCHEIFLINGER, F.
;   APPLICANT:  FALKNER, F. G.
;   TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;   NUMBER OF SEQUENCES:  52
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Foley & Lardner
;     STREET:  1800 Diagonal Road, Suite 500
;     CITY:  Alexandria
;     STATE:  VA
;     COUNTRY:  USA
;     ZIP:  22313-0299
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0,  Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/232,463
;     FILING DATE:
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US/07/935,313
;     FILING DATE:  26-AUG-1991
;     APPLICATION NUMBER:  EP 91 114 300.6
;     FILING DATE:  26-AUG-1991
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  BENT, Stephen A.
;     REGISTRATION NUMBER:  29,768
;     REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (703)836-9300
;     TELEFAX:  (703)683-4109
;     TELEX:  899149
; INFORMATION FOR SEQ ID NO:  14:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  7218 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   IMMEDIATE SOURCE:
;     CLONE:  pT2gpt-F1s
US-08-232-463-14
```

```
Query Match     2.0%;    Score 63.8;  DB 1;  Length 7218;
Best Local Similarity   10.9%;  Pred. No. 9.1e-08;
Matches    53;  Conservative  226;  Mismatches  208;  Indels  208;  Gaps    0;

Qy    843 TATTACTAACATATTAAACGTTTATCTTTGAGTTCGTCGTGTTTTAGGTTTCATCATGAG 902
          | ||| |||||
Db   1551 TGTGACTAGCGTAGCATCGCTTCTAGACGCATCTATTCAGTTTCAAAAAACGGCATGTAG 1492

Qy    903 TAACTACAGTGGAAAATCCTCTATGGACCCTGATTATAATGTGGATGAAGCTAAGTCCTG 962
          | |||| || ||||
Db   1491 GCATCACTGTAATTACCTATCCTATGCAAGTAGTGAAGATAGAAGAATTTGGTACRRR 1432

Qy    963 GTCCACTAGACCGGAGTGAGAGCAACATGTTTACGAGAGCTATAGGGATGAATTTGAACG 1022
          :: ::: :::: :::
Db   1431 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1372

Qy   1023 CTCTGCAGCTCGACGTAATCAAAGAAGAGAGCTGAAATCGCTAGAGGAAAGAGGGCCGATGTC 1082
          ::::::::: :::::: :::
Db   1371 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1312

Qy   1083 GAGTAGATATGAGCTGATTGATGAGATATCAAAACTGAGTATGAGCCAGAGTCATGGCG 1142
          ::: :::: ::::: :::::   :::: :::
Db   1311 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1252

Qy   1143 CAAGGAGACGAAGCTACTGAACAAATCCGACGAGTTACAGTGGAGGAGTATATCAGATT 1202
          ::::: :::: ::::: :::::   ::: :::
Db   1251 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1192

Qy   1203 CTTTGAGATGAATGACTTCTGGGGAACGAGGTATCCCTGATATGAGACTTTAGCCCAGTT 1262
          ::: ::::::::: :::
Db   1191 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1132

Qy   1263 GGGGTTACTGGAGGACGTGCGAGCATCTGTTCGAGAAGTGTCATCTGATAAGGAGGAGACA 1322
          ::::  :::::::: :: :::::::::
Db   1131 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1072

Qy   1323 ATCGAGT 1329
          ::  :::
Db   1071 RRRRRAT 1065
```

RESULT 6
US-08-658-136-2
; Sequence 2, Application US/08658136
; Patent No. 6071717
GENERAL INFORMATION:
APPLICANT: KLINGER, KATHERINE W
APPLICANT: LANDES, GREGORY M
APPLICANT: BURN, TIMOTHY C
APPLICANT: CONNORS, TIMOTHY D
APPLICANT: DACKOWSKI, WILLIAM
APPLICANT: GERMINO, GREGORY
APPLICANT: QIAN, FENG
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
  ADDRESSEE: GENZYME CORPORATION
  STREET: ONE MOUNTAIN ROAD
  CITY: FRAMINGHAM
  STATE: MASSACHUSETTS
  COUNTRY: USA
  ZIP: 01701
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/658,136
  FILING DATE:
  CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
  NAME: LASSEN, ELIZABETH

REGISTRATION NUMBER: 31,845
REFERENCE/DOCKET NUMBER: GEN4-17.8
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 508-872-8400
  TELEFAX: 508-872-5415
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
  LENGTH: 53526 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  MOLECULE TYPE: DNA (genomic)
US-08-658-136-2

```
Query Match      1.8%;    Score 55;  DB 3;  Length 53526;
Best Local Similarity   46.0%;  Pred. No. 0.00013;
Matches   225;  Conservative    0;  Mismatches  260;  Indels    4;  Gaps    1;

Qy     96 CTCTCCACTTACTCGGCTTCATCGCCATCTCTCAACATACTCGACCTCGC 155
          ||||||||
Db  35885 CTCTCTCCCCCTCTTTTCTCCCTTCCCCTCCCCTCCCCTCTTCCCC 35944

Qy    156 GATATCACTCGAGCTCGCGCGTTCTCACCGGCTCTCCATCGTCACCGCCTGCTCCCTCTC 215
          ||||||||||
Db  35945 TCCCCTCTCTTCCCCCTCTTCCCCTCTCCCCTCTCCCCTCTTCCCCTCTT 36004

Qy    216 TCCAAGGAAACAACTCGAGCTCCATTCACTCGACCTCACTCACCAAGCCGGCT 275
          || ||
Db  36005 TCCCTCCCCCTCTCCCCTCTCCCCTCTTCCCCTCTTCCCCTCTTCCCCTCCCC 36064

Qy    276 TCACCACTTCTAGCTCTTAACCACTCGACCACCTTCACCATCAACCAATCAAATCGTTTT 335
          ||||||||
Db  36065 TTCCCCTCTCCCCTCTTCCCCTCTCCCCTCTCCCCTTCCCCCTC 36124

Qy    336 CTCCTCCATTAAAGTTGACATACTCGACGCGTGAACACTTATCACCTTCAAGCTCCTCA 395
          ||||||||
Db  36125 CTCTTCTTCCCCTCTCTTCCCCTCTCCCCTCTCCCCTCTCCCCTTCTCT 36184

Qy    396 TCTTCTTCATCGTTTCCAACACGCTGCTCTCATCCCCCAGAAAGTTGTCATCACCTCT 455
          ||||||||
Db  36185 TCTCCCCTCCCTTCTCTTCCCCTCTCTTCCCCTCTTCTTCCCTGCCCTTCCT 36244

Qy    456 CACTCATCACCAGTTCACTCGATTCAGCAACAACCAAACTCGACCTCGTCTCCTCCTGCCACTC 515
          ||||||||
Db  36245 CTTCCCCGTTCCCCCTCTCCCCTCTCCTCTCCTCTCCCT----CCCCTC 36300

Qy    516 ATAGTCACTCGATCTCTCCTCACCATCTTCATCATCTCCCTCTGTC 575
          ||||||||
Db  36301 TTCCTCTTTCCCTTCCTCCCTCTCCCCTTCCCCTCCCCTCTCCCCT 36360

Qy    576 TCGCTCCAC 584
          |||||||
Db  36361 CCGCTTCCC 36369
```

RESULT 7
US-08-658-136-1
; Sequence 1, Application US/08658136
; Patent No. 6071717
GENERAL INFORMATION:
APPLICANT: KLINGER, KATHERINE W
APPLICANT: LANDES, GREGORY M
APPLICANT: BURN, TIMOTHY C
APPLICANT: CONNORS, TIMOTHY D
APPLICANT: DACKOWSKI, WILLIAM
APPLICANT: GERMINO, GREGORY
APPLICANT: QIAN, FENG
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
  ADDRESSEE: GENZYME CORPORATION
  STREET: ONE MOUNTAIN ROAD
  CITY: FRAMINGHAM

```
;     STATE: MASSACHUSETTS
;     COUNTRY: USA
;     ZIP: 01701
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION.NUMBER: US/08/658,136
;     FILING DATE:
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: LASSEN, ELIZABETH
;     REGISTRATION NUMBER: 31,845
;     REFERENCE/DOCKET NUMBER: GEN4-17.8
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 508-872-8400
;     TELEFAX: 508-872-5415
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 53577 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-08-658-136-1

Query Match        1.8%;    Score 55;   DB 3;   Length 53577;
Best Local Similarity  46.0%;  Pred. No. 0.00013;  Indels     4;  Gaps     1;
Matches    225;    Conservative    0;  Mismatches    260;  Indels    4;  Gaps    1;

Qy      96 CTCTCTCACTTACTCGGCTTCATCGCTCTCATCGCCATCTCAACATACTCGACCTCGC 155
           ||||||||||||||| ||||||||||||||||||||||| |||||||||||||||||||
Db   35880 CTCTCTCCCCCTTCTTTCCCTCTTCCCCTCTCCCCTCTCCCCTCCCTCTCCTCTTCCCC 35939

Qy     156 GATATCACTCGAGCTGCGCGCGTTCTCACCGCTCTCCATCGTCACCGCTCCCTCTC 215
           |||||||||||||||||||||||||||||||||||||| |||||
Db   35940 TCCCCTCCTCTTCCCCTCCCCTCCCTCTTCCCCTCTCCCTCTCCCCTCCCTCTCTT 35999

Qy     216 TCCAAGGAAACAACTCGAGCTCTCCATTCACTCACTCGACCTCTACCACCCAAGCCGGCT 275
           ||||  |||||||||||||||||||||| |||||
Db   36000 TCCCTCCCCCTCTTCCTCCCCTCTCCCCCTCTTCCCCCCTCTTCCCCTCTTCCCCCC 36059

Qy     276 TCACCACTTCTAGCTCTTAACCACTCGACCACCTTCACCATCATCAACCAATCAAATCGTTTT 335
           |||||||||||| |||||||||||||
Db   36060 TTCCCCTCCCCCTCCTCTTCCCCTCCCCCTCCTCTTCCCCTCCCCCTCCCTCCCCTC 36119

Qy     336 CTCTCCCATTAAAGCTTGACATACTGACGCCGCTGAACACTTATCACCTTCAAGCTCCTCA 395
           ||||||||||||| |||||||||||
Db   36120 TCCTTCCTTCCTCTCCCTTCCCCTCCCCTCCTCTTCCCCTCCCCCTCCCCTTTCTCT 36179

Qy     396 TCTCTTCATCGTTTTCCAAACACCGCTGCTTCATCCCCCACGAAAAGCTTGTCATCACCTCT 455
           ||||||||||||| ||||||||||||
Db   36180 TCTCCTTCCCCTTCTCTTCCCCTCTTTTCTTCCCCTGTCTTCCCTGCCCCTCCT 36239

Qy     456 CACTCATCACCAGTTCACTCGATTCAGCAACCAAACTCGACCTCGTTCTCTCTTGCCACTC 515
           |||||||||||||||||||||||||||||||||||||||| ||||
Db   36240 CTTCCCCTCCCCTCTCCCCTCCCCTCCCTCTTCCCCTCTCCCT----CCCCTC 36295

Qy     516 ATAGTCACTCGATCTCTCCACCATCTTCATCATCTCCCTTACTCGACCACGTGCGTC 575
           |||||||||||||||||||||||||||||||||||||||||||||||
Db   36296 TTCCCCTCTTTCCTCTTCCTCCCCTCTCCTCCCCTCCCTTCCCCTTCCCCTCCCCCT 36355

Qy     576 TCGCTCCAC 584
           ||||||
Db   36356 CCGCTTCCC 36364
```

RESULT 8
US-10-204-708-25
; Sequence 25, Application US/10204708
; Patent No. 6677731

```
; GENERAL INFORMATION:
;     APPLICANT: OLEK, Alexander
;     APPLICANT: PIEPENBROCK, Christian
;     APPLICANT: BERLIN, Kurt
;     TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;     TITLE OF INVENTION: by Assessing DNA Methylation
;     FILE REFERENCE: 5013.1012
;     CURRENT APPLICATION NUMBER: US/10/204,708
;     CURRENT FILING DATE: 2003-05-06
;     PRIOR APPLICATION NUMBER: PCT/EP01/03971
;     PRIOR FILING DATE: 2001-04-06
;     PRIOR APPLICATION NUMBER: DE 10019058.8
;     PRIOR FILING DATE: 2000-04-06
;     PRIOR APPLICATION NUMBER: DE 10019173.8
;     PRIOR FILING DATE: 2000-04-07
;     PRIOR APPLICATION NUMBER: DE 10032529.7
;     PRIOR FILING DATE: 2000-06-30
;     PRIOR APPLICATION NUMBER: DE 10043826.1
;     PRIOR FILING DATE: 2000-09-01
;     NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 25
;     LENGTH: 6583
;     TYPE: DNA
;     ORGANISM: Artificial Sequence
;   FEATURE:
;     OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-25

Query Match        1.5%;    Score 45.4;   DB 4;   Length 6583;
Best Local Similarity  49.8%;  Pred. No. 0.022;
Matches    141;    Conservative    0;  Mismatches    141;  Indels    1;  Gaps    1;

Qy    2771 AGGTAAGCGGCTCACTTCACCATTATTATATATATCATCTCTTGTGATTTGTTCTTTATTTT 2830
           ||||   |||||||||| |||  ||||||||||||||||||| ||||||||| ||||
Db    5792 AAGTATAGGATTATTTTGTATTTTTTTTTTTTTTTTGAATTTTTTTTTTTTGTTTTTAATTGAAATA 5851

Qy    2831 GTTTCAGTGATTGATTTGTCCTGAGTACTCTCTCCAAGTTTATTCACACAGTGGACTG 2890
           ||||||| ||||| ||||| ||
Db    5852 TTTATTTTTTATAAATTGTTTGAGTTTTATTTTTATTTTGTACGTTGATTTT 5911

Qy    2891 TGTGATTTAAGTTGGGGGAGGGCTCAGGAAGTAGTGTGCATTGCATTGTATGTATATATTTTTAAGTC 2950
           |||| |||||||||||||||||||
Db    5912 TATGATTTAAGTTTTTTGTAGAGGTAATGCGATGTTATCGTATGTATGGTTTTTGTTTGTG 5971

Qy    2951 TGCATTCATCTAAGGCATAGAAAAACCAAAAAAATTAAAAATTCAGAAAAATGATTTC 3010
           |||||||||||||||
Db    5972 TTTGGTTTTTGAAGTAT-GTGTTAGGAGTGAAAGTGTAAAGTATTTTGGAAAGATTAA 6030

Qy    3011 ACAAAAAAGAGTGTTCATGTAGTTGCATTACATTTAGGATCA 3053
           ||||||||| |||||
Db    6031 ATAAATAATAAGTAATAGTTAAGTGGTATTAGAGGTAGTATTA 6073
```

RESULT 9
US-09-165-264-12/c
; Sequence 12, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
;   APPLICANT: Vinayagamoorthy, Thuraiayah
;   TITLE OF INVENTION: Multi-Loci Genomic Analysis
;   FILE REFERENCE: 44747
;   CURRENT APPLICATION NUMBER: US/09/165,264
;   CURRENT FILING DATE: 1998-10-01
;   NUMBER OF SEQ ID NOS: 14
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
;   LENGTH: 318
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-12

Query Match   1.4%;   Score 42.6; DB 3; Length 318;
Best Local Similarity   47.0%;   Pred. No. 0.02;
Matches 132; Conservative 0; Mismatches 149; Indels 0; Gaps   0;

```
Qy     41 AGACACACAAGCGCCATCGCTTTCCCTCACAACTTCACTCGACACCGCCCGCTCTC 100
             ||||||||       |||||||||||||||||||
Db    311 ACACACCAACCAACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 252

Qy    101 TCACTTACTCGGCTTCATCGCTCTCAACATACTCTCAACATACTCGACCTCGCGATAT 160
                                                       ||||||||
Db    251 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 192

Qy    161 CACTCGAGCTCGCGCGTTCTCCACCGCCTCTCCATCGTCACCGGCTGTCCCTCTCCAA 220
                       ||||||           |||||        |||||
Db    191 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 132

Qy    221 GGAAACAACTCGAGCTCTCCATTTCACTCGACTCTCACCACCAAGCCGGCTTCACC 280
                        |||||||               |||||
Db    131 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 72

Qy    281 ACTTCTAGCTCTTAACCACTGACCACCTTCACCATCAACC 321
                   |||||
Db     71 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 31
```

RESULT 10
US-08-770-379-20/c
; Sequence 20, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
;   APPLICANT: Chang, Yuan
;   APPLICANT: Bohenzky, Roy A.
;   APPLICANT: Russo, James J.
;   APPLICANT: Edelman, Isidore S.
;   APPLICANT: Moore, Patrick S.
;   TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
;   TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
;   NUMBER OF SEQUENCES: 20
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Cooper & Dunham LLP
;   STREET: 1185 Avenue of the Americas
;   CITY: New York
;   STATE: New York
;   COUNTRY: U.S.A.
;   ZIP: 10036
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/770,379
;   FILING DATE:
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: White, John P.
;   REGISTRATION NUMBER: 28,678
;   REFERENCE/DOCKET NUMBER: 52342
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 278-0400
;   TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 20:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 32207 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-08-770-379-20

Query Match   1.3%;   Score 41.6; DB 2; Length 32207;
Best Local Similarity   44.7%;   Pred. No. 0.82;
Matches 161; Conservative 0; Mismatches 199; Indels 0; Gaps   0;

```
Qy    121 CTCTCATCGCATCTCTCAACATATCGACCTCGCGATATCACTCGAGCTCGCCGCTTCT 180
             |||||     ||||       |||||          |||||
Db  32202 CTCCCCCCCTTCACCCCTTCACCCCTTCTCCAAGGAAACAACTCGAGCTCC 32143

Qy    181 CACCGCCTCCGTCACCGCGTCGCTGCTCTCTCCAAGGAAACAACTCGAGCTCC 240
                     ||||          |||||
Db  32142 CCCCCATCCCCCTTTCCCCCATCCCCCATCCCCCTTTCCCCCCC 32083

Qy    241 ATTTCACTCACTCGACCTCTCACCACCAAGCCGGCTTCACCACTTCTAGCTCTTAACCACT 300
             |||          ||||||        |||||          ||||
Db  32082 ATCCCCCCTTTCCCCCATCCCCCGTTCCCCCATCCCCCGTTTCCCCCCATTCCCCCT 32023

Qy    301 CGACACCTTCACCATCAACCAATCAACCAAATGTTTTCTCCTCCATTAAAGCTTGACATACT 360
             ||          |||||               |||||
Db  32022 TTCCCCCATCCCCCTTTCCCCCATCCCCCGTTTCCCCCATCCCCCTTTCCCCCCA 31963

Qy    361 CGACCGCTGAACACTTATCACCTTCAAGCTCCTCATCGTCGTTCCAACACCGCT 420
             ||          |||||        |||||          |||
Db  31962 TCCCCCTTTCCCCCATTCCCCCTTTCCCCCATCCCCCGTTTCCCCCATCCCCCTT 31903

Qy    421 GCTCTCATCCCCACGAAAGCTTGTCATCACCCCCTTTCCATCATCACCAGTTCACTCGATTC 480
             ||||          ||||          |||||          ||||
Db  31902 TCCCCCCATTCCCCCCATCCCCCATCACCCCCTTTCCCCACATCCCCCATCACCCCCATCC 31843
```

RESULT 11
US-08-757-669A-20/c
; Sequence 20, Application US/08757669A
; Patent No. 6183751
; GENERAL INFORMATION:
;   APPLICANT: Chang, Yuan
;   APPLICANT: Bohenzky, Roy A.
;   APPLICANT: Russo, James J.
;   APPLICANT: Edelman, Isidore S.
;   APPLICANT: Moore, Patrick S.
;   TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
;   TITLE OF INVENTION: SEQUENCES AND USES THEREOF
;   NUMBER OF SEQUENCES: 20
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Cooper & Dunham LLP
;   STREET: 1185 Avenue of the Americas
;   CITY: New York
;   STATE: New York
;   COUNTRY: U.S.A.
;   ZIP: 10036
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/757,669A
;   FILING DATE:
;   CLASSIFICATION: 424
;   ATTORNEY/AGENT INFORMATION:
;   NAME: White, John P.
;   REGISTRATION NUMBER: 28,678
;   REFERENCE/DOCKET NUMBER: 45185-F
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 278-0400
;   TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 20:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 32207 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-08-757-669A-20

Query Match   1.3%;   Score 41.6; DB 3; Length 32207;
Best Local Similarity   44.7%;   Pred. No. 0.82;

Matches 161; Conservative   0; Mismatches 199; Indels   0; Gaps   0;

Qy   121 CTCTCATCGCCATCTCTCAACATACTCGACCTCGGGATATCACTCGAGCTCGCCGCTTCT 180
                ||||||||||||||| || | | ||| | || || ||||||
Db 32202 CTCCCCCCTTTCCCCCCATTCCCCTTTCCCCCCCCCCCCCCCCCCCCCCCCCCTTTC 32143

Qy   181 CACCGGCTCTCCATCGTCACCGCCTTCCCCTTTCCCCAAGGAAACAACTCGAGCTCTCC 240
              |||||||| | |||||||||||||||| |||||||||||||||||||||
Db 32142 CCCCCCATTCCCCTTTCCCCCCATCCCCTTTCCCCCCCC 32083

Qy   241 ATTTCACTCACTGACCTCTACCACCAAGCCGGCTTCACCACTTCTAGTCTTAACCACT 300
             | | | |||||||||||||| |||||||||||||||||||||||||||
Db 32082 ATCCCCCCTTTCCCCCATCCCCGTTTCCCCCATCCCCTTTCCCCCCATTCCCCT 32023

Qy   301 CGACCACCTTCACCATCAACCAACCAATCAAATCGTTTCTCCTTCCATTAAAGCTTGACATACT 360
             | | | || |||||||||||||||||||| | | | | || | ||||||
Db 32022 TTCCCCCATCCCCCTTTCCCCCATTCCCCTTTCCCCCATTCCCCCCA 31963

Qy   361 CGACCGCTGAACACTTATCACCTTCAAGCTCTCATCTCTTCAGCTCCTCATTCCCCTT 420
             |||||||||| |||| | ||| | |||||| |||||||||||||||||
Db 31962 TCCCCCCTTTCCCCCATCCCCTTTCCCCCATTCCCCTTTCCCCCCATTCCCCTT 31903

Qy   421 GCTCTCATCCCCCAGCAAAGCTTGTCATCACCTCTCACTCACCCCTTTCCACTCACTGATTC 480
             |||| | || || |||||||||| || |||||||| | | ||||
Db 31902 TCCCCCCATTCCCCATCCCCTTTCCCACCATTCCCCATCACCCCTTTCCCCATTCCCCATCC 31843

RESULT 12
US-09-230-371A-20/c
; Sequence 20, Application US/09230371A
; Patent No. 6348586
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A
; APPLICANT: Russo, James J
; APPLICANT: Edelman, Isidore S
; APPLICANT: Moore, Patrick S
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 45185-G-PCT-US
; CURRENT APPLICATION NUMBER: US/09/230,371A
; CURRENT FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: PCT/US97/13346
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 20
; LENGTH: 32207
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-230-371A-20

Query Match   1.3%; Score 41.6; DB 4; Length 32207;
Best Local Similarity   44.7%; Pred. No. 0.82;
Matches 161; Conservative   0; Mismatches 199; Indels   0; Gaps   0;

Qy   121 CTCTCATCGCCATCTCTCAACATACTCGACCTCGGGATATCACTCGAGCTCGCCGCTTCT 180
                ||||||||||||||| || | | ||| | || || ||||||
Db 32202 CTCCCCCCTTTCCCCCCATTCCCCTTTCCCCCCCCCCCCCCCCCCCCCCCCCCTTTC 32143

Qy   181 CACCGGCTCTCCATCGTCACCGCCTTCCCCTTTCCCCAAGGAAACAACTCGAGCTCTCC 240
              |||||||| | |||||||||||||||| |||||||||||||||||||||
Db 32142 CCCCCCATTCCCCTTTCCCCCCATCCCCTTTCCCCCCCC 32083

Qy   241 ATTTCACTCACTGACCTCTACCACCAAGCCGGCTTCACCACTTCTAGTCTTAACCACT 300
             | | | |||||||||||||| |||||||||||||||||||||||||||
Db 32082 ATCCCCCCTTTCCCCCATCCCCGTTTCCCCCATCCCCTTTCCCCCCATTCCCCT 32023

Qy   301 CGACCACCTTCACCATCAACCAATCAAATCGTTTCTCCTTCCATTAAAGCTTGACATACT 360
             | | | || |||||||||||||||||| | | | | || | ||||||
Db 32022 TTCCCCCATCCCCCTTTCCCCCATTCCCCTTTCCCCCATTCCCCCCA 31963

Qy   361 CGACCGCTGAACACTTATCACCTTCAAGCTCTCATCTCTTCAGCTCCTCATTCCCCTT 420

RESULT 13
US-09-621-976-2813/c
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
;
; SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813

Query Match   1.3%; Score 40.6; DB 4; Length 832;
Best Local Similarity   10.9%; Pred. No. 0.15;
Matches 28; Conservative 125; Mismatches 104; Indels   0; Gaps   0;

Qy   2762 ATTCGACTGAGGTAAGGCGCCTCACTTCACCATTATATATATCATCTCTTGTGATTTGTT 2821
             ::|| : : : : ||: :: || :: ||:|:| ||:|::|: :::|::| :::
Db   281 WYTMARGRMWSYAWGKWKSMRSAMSMCTRMYYKKGSTYWTMKCTCATWCYWYWKRMW 222

Qy   2822 CTTTATTTTGTTTCAGTGATTGGATTGTCCTGGAGTACTCTCTTCCAAGTTTATTCACAC 2881
             ::|| | || :: |::|: :: || : : ::||||: :::
Db   221 SKTCWSGSRGGYMTSYTSTRSYSMYWASWMYTMCWWWGRWWSTYWYMAWGKKWWRYATTW 162

Qy   2882 AGTGGACTGTGTGATTTAAGTTTGGGGAGGGCTCAGGAAGTATGTTGCATTGTATATAT 2941
             ::| : : : : :: | ||: : : :: : : :::|::||:: ::| :|::|:
Db   161 RRAMMWWWAAAWTMWWYMWWAWCMSSRGAAMYRRTMMMWGYRYWWRKKSYRRTRCAWAYAW 102

Qy   2942 TTTTAAGTCTGCATTCATCTAAGGCATAGAAAAAACCAAAAAAAATTAAAAATTTCAGAA 3001
             :|:.. ::: : : :: :| | : :::| :::| | ::: :::| :||::|:|
Db   101 KTKRSYYWCWRWKWKRCMMMMMAMAYGKTMMMRACWKTRYWRWWAWAMWRMWTMMMMY 42

Qy   3002 AATGATTTCACAAAAAA 3018
             :: : :|| :: ::
Db   41 YWYWRAMKRRWWRKWR 25

RESULT 14
US-09-257-179-22
; Sequence 22, Application US/09257179
; Patent No. 6410709
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 29 Human Secreted Proteins
; FILE REFERENCE: PZ015P1
; CURRENT APPLICATION NUMBER: US/09/257,179
; CURRENT FILING DATE: 1999-02-25
; EARLIER APPLICATION NUMBER: PCT/US98/17709
; EARLIER FILING DATE: 1998-08-27
; EARLIER APPLICATION NUMBER: 60/056,270
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,271
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,247

```
;   EARLIER FILING DATE: 1997-08-29
;   EARLIER APPLICATION NUMBER: 60/056,073
;   EARLIER FILING DATE: 1997-08-29
;   NUMBER OF SEQ ID NOS: 128
;   SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 22
;   LENGTH: 487
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-257-179-22

Query Match      1.3%;   Score 40.4;  DB 4;  Length 487;
Best Local Similarity    47.9%;   Pred. No. 0.12;
Matches  116;  Conservative   0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy    2780 CCTCACTTCACCATTATTATTATATCATCTCTTGTGATTTGTTCTTTATTTGTTTCAGTG 2839
                ||||||  ||||||| ||||      || |||| ||||   ||| |   |||||
Db     226 CCTTTTTTGCCAGATGATGATTTTTTTATTTTATTTTGTACTTTACTGGATGACTGTG 285

Qy    2840 ATTGGATTTGTCCTGAGTACTCTCTTCCAAGTTTATTCACACAGTGGACTGTGTGATTTA 2899
                ||||      || |||         ||| |  ||| |  ||  | |   |||
Db     286 AAGCGATGAGTATTGGGGTTGGGTAGGTGTGTTGATTTTGAGAGTGCATGTTAAGAACTG 345

Qy    2900 AGTTTGGGGGAGGGCTCAGGAAGTATGTTGCATTGTATATATTTTAAGTCTGCATTCAT 2959
                 ||| |  | | |       |||| || ||       | ||  ||
Db     346 AAGGGAACTACTTGAGATGACTTAAGAAGCATCCCATGCAAATATCTTGTTTTGCCCTA 405

Qy    2960 CTAAGGCATAGAAAAACCAAAAAAAATTAAAAATTTCAGAAAAATGATTTCACAAAAAA 3019
                    |||      |||||||||        |    |      |  |||||||||
Db     406 ATAAATATTCAGAAAGATAAAAGATAAAAAAAAAAAAAAAAA 465

Qy    3020 GA 3021
                ||
Db     466 AA 467


RESULT 15
US-09-165-264-7/c
;   Sequence 7, Application US/09165264
;   Patent No. 6197510
;   GENERAL INFORMATION:
;   APPLICANT: Vinayagamoorthy, Thuraiayah
;   TITLE OF INVENTION: Multi-Loci Genomic Analysis
;   FILE REFERENCE: 44747
;   CURRENT APPLICATION NUMBER: US/09/165,264
;   CURRENT FILING DATE: 1998-10-01
;   NUMBER OF SEQ ID NOS:  14
;   SOFTWARE: PatentIn Ver. 2.1
;   SEQ ID NO 7
;   LENGTH: 320
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-7

Query Match      1.3%;   Score 40.2;  DB 3;  Length 320;
Best Local Similarity    46.3%;   Pred. No. 0.1;
Matches  132;  Conservative   0;  Mismatches  153;  Indels    0;  Gaps    0;

Qy      37 CAAAAGACACAACAAGCCGCATCGCCATCCTTCCCTCACAACTCTCACTCGACCACCGCCCGC 96
                | || || | ||||  | | | |   ||  |   || ||| |||||||||
Db     319 CTAAGGGCATGCTAGCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 260

Qy      97 TCTCTCACTTACTCGGCTTCATCGCTCTCATCGCCATCTCAACATACTCGACCTTCGCG 156
                |   |  |  |    || ||||||||||||||||||||||||||||
Db     259 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 200

Qy     157 ATATCACTCGAGCTGGCGCTTCTCACCGCCTCTGCTCCATCGTCACCGCCTGCTCCCTCT 216
                 ||||||||||| ||||    |       |||  |  ||
Db     199 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 140

Qy     217 CCAGGAAACAACTCGAGCTCCATTTCACTCACTCGACCTCTACCACCAAGCCGGCTT 276
```

(right column)

```
Db     139 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 80
                ||  |||||||||||||||||||||||||||||||||||||||| ||  |||
Qy     277 CACCACTTCTAGCTCTTAACCACTCGACCACTTCACCATCAACC 321
                |||  |||||||||||||||||||||| ||  ||||||
Db      79 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 35


Search completed: September  9, 2004, 09:17:09
Job time : 243 secs
```

OM nucleic - nucleic search, using sw model

```
          GenCore version 5.1.6
          Copyright (c) 1993 - 2004  Compugen Ltd.
```

Run on: September 9, 2004, 03:26:27 ; Search time 7044 Seconds
                              (without alignments)
                    13235.339 Million cell updates/sec

Title: US-10-088-384A-27
Perfect score: 3122
Sequence: 1 ttcatatattcgacctcttc........atgagatagccttgtaagca 3122

Scoring table: IDENTITY NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 1493109276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

```
Database :     EST:*
            1: em_estba:*
            2: em_esthum:*
            3: em_estin:*
            4: em_estmu:*
            5: em_estcv:*
            6: em_estpl:*
            7: em_estro:*
            8: em_htc:*
            9: gb_est1:*
           10: gb_est2:*
           11: gb_htc:*
           12: gb_est3:*
           13: gb_est4:*
           14: gb_est5:*
           15: em_estfun:*
           16: em_estom:*
           17: em_gss_hum:*
           18: em_gss_inv:*
           19: em_gss_pln:*
           20: em_gss_vrt:*
           21: em_gss_fun:*
           22: em_gss_mam:*
           23: em_gss_mus:*
           24: em_gss_pro:*
           25: em_gss_rod:*
           26: em_gss_phg:*
           27: em_gss_vrl:*
           28: gb_gss1:*
           29: gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                    SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	725.6	23.2	855	28	BH237235	BH237235 AUIIB10TR
c	2	684.4	21.9	809	28	BH237437	BH237437 AUIIB84TR
	3	682.2	21.9	801	28	BH237501	BH237501 AUIIB05TR
	4	682.2	21.9	817	28	AQ958311	AQ958311 LERAW77TR
	5	663.4	21.2	781	28	BH244328	BH244328 AUJAA71TF
c	6	647.2	20.7	848	28	BH237552	BH237552 AUIIB35TF
	7	622.6	19.9	813	28	BH244305	BH244305 AUJAA72TF
c	8	611.6	19.6	806	28	BH235546	BH235546 AUIIA07TF
c	9	604.6	19.4	681	28	BH244756	BH244756 AUIJA07TF
	10	601	19.3	691	28	BH244796	BH244796 AUIJA58TR
	11	598.4	19.2	1101	28	B08362	B08362 T24I6-T7.1
c	12	587.4	18.8	704	28	BH237544	BH237544 AUIIA17TF
c	13	585.6	18.8	792	28	BH237151	BH237151 AUIIB06TR
	14	577.2	18.5	670	28	BH237213	BH237213 AUIIB28TR
	15	576.2	18.5	756	28	BH244293	BH244293 AUJAA06TF
c	16	572.4	18.3	725	28	BH244787	BH244787 AUIJA90TF
c	17	568.2	18.2	713	28	BH237198	BH237198 AUIIA50TR
	18	553.6	17.7	684	28	BH244310	BH244310 AUJAA82TF
	19	553.4	17.7	663	28	BH235487	BH235487 AUIV869TR
c	20	551.4	17.7	662	28	BH235467	BH235467 AUIVB93TF
c	21	550.6	17.6	637	28	BH235552	BH235552 AUIVA84TF
c	22	543.4	17.4	685	28	BH235356	BH235356 AUIVB65TR
	23	541.2	17.3	645	28	BH235469	BH235469 AUIVA83TR
	24	529	16.9	617	28	AQ958677	AQ958677 LERA206TR
	25	528	16.9	740	28	BH235520	BH235520 AUIVC93TF
	26	526.6	16.9	682	28	BH244766	BH244766 AUIJA10TF
	27	521.2	16.7	582	28	B77880	B77880 T29H16TF TA
	28	519.6	16.6	771	28	BH237289	BH237289 AUIIB53TR
	29	509.8	16.3	630	28	BH235434	BH235434 AUIVA43TF
	30	504.2	16.1	558	28	BH244805	BH244805 AUIJA85TF
c	31	499	16.0	571	28	BH235416	BH235416 AUIVB87TR
c	32	499	16.0	623	28	BH235383	BH235383 AUIVB76TR
	33	498.2	16.0	621	28	BH244296	BH244296 AUJAA87TF
c	34	498	15.9	554	28	BH244834	BH244834 AUIJA43TR
	35	497.2	15.9	638	28	BH244311	BH244311 AUJAA74TF
	36	487.4	15.6	545	28	B67818	B67818 T24I16TF TA
	37	486.6	15.6	627	28	BH244334	BH244334 AUJAA04TF
	38	486	15.6	650	28	BH235456	BH235456 AUIVC07TR
	39	485.2	15.5	765	28	BH244841	BH244841 AUIJA22TF
	40	484.4	15.5	638	28	BH235402	BH235402 AUIVB39TR
	41	481	15.4	645	28	BH235572	BH235572 AUIVB57TF
c	42	477.4	15.3	723	28	BH243461	BH243461 AUIRA66TR
c	43	476.6	15.3	638	28	BH235471	BH235471 AUIVB32TF
c	44	476.2	15.3	583	28	BH235549	BH235549 AUIVC80TR
c	45	475	15.2	654	28	BH235355	BH235355 AUIVC22TF

                              ALIGNMENTS

```
RESULT 1
BH237235/c
LOCUS       BH237235                 855 bp    DNA    linear    GSS 13-NOV-2001
DEFINITION  AUIIB10TR AUII Arabidopsis thaliana genomic clone AUIIB10, genomic
            survey sequence.
ACCESSION   BH237235
VERSION     BH237235.1  GI:16907606
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 855)
  AUTHORS   Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Utterback,T.V.,
            Feldblyum,T.V. and Fraser,C.M.
  TITLE     Survey sequencing of Arabidopsis thaliana BAC T2P24
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: AUIIB10TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            From Wash. U contig 849.
            Seq primer: TR
```

```
Qy    2334  CTCAGGGCTATCAAGT 2349
            ||||||||||||||||
Db      16  CTCAGGGCTATCAAGT 1

RESULT 2
BH237437/c
LOCUS       BH237437              809 bp    DNA     linear   GSS 13-NOV-2001
DEFINITION  AUIIB84TR AUII Arabidopsis thaliana genomic clone AUIIB84, genomic
            survey sequence.
ACCESSION   BH237437
VERSION     BH237437.1  GI:16907915
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 809)
  AUTHORS   Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Utterback,T.V.,
            Feldblyum,T.V. and Fraser,C.M.
  TITLE     Survey sequencing of Arabidopsis thaliana BAC T2P24
  JOURNAL   Unpublished (2001)
  COMMENT   Other_GSS: AUIIB84TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            From Wash. U contig 849.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..809
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="AUIIB84"
                     /clone_lib="AUII"
                     /note="Vector: pHOS2; Site 1: BstXI; 2-3 kb sheared BAC
                     DNA inserted into pHOS2 using BstXI linkers"
ORIGIN
Query Match      21.9%;  Score 684.4; DB 28; Length 809;
Best Local Similarity 91.1%;  Pred. No. 1.6e-114;
Matches 738;  Conservative 0; Mismatches 71; Indels 0; Gaps 1;

Qy  1603  GCTACTTTCAGCGCTCGGTTGCCAATGTTTTTTACTCACGGGAGTCTACAGGCACCGTGT 1662
          ||||||||||||||||||||||||||||||  |||| ||||||||||||||||||||||||
Db   809  GCTACTTTCAGCGCTCGGTTGCCAACG-TCTTTACTCACGGGAGTCTACAGACACCGTGT 751

Qy  1663  CTAACACAGACATGAAGATGATAGATCCAGCGCTTATAGGGATTCTCCGCCTTACAAAAG 1722
          ||||||||||||||||||||||||||||||||||||||||| ||||||||  ||||||||
Db   750  CTAACACGACACATGAAGATGATAGATCCAGCGCTTATAGGGATTCTCCGCCGGACGAAGG 691

Qy  1723  GAAAGAAATGTCCTGAGAGAGATCTTAACGACTCACCAGTAATGCCTCGTTGATCC 1782
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   690  GGAAGAAATGTCTTGAGAGAGGCGATCTCAACACGCCACCAGTAATGCCTCGTTGATCC 631

Qy  1783  ATCTGTGTGGTACATGAAGTGGGGCGTTGACGAAGAGGTAAGAGGAGCAC 1842
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   630  ATCTGTGTGATACAGGACGTTGACGAAGGGGCGTTGACGAAGAGGTGAGAGGAGCAC 571

Qy  1843  TATGCGTGGGTGGCCGTTGTGCACGCAATTCTGAAAGTTTGTGGAGTTCCGTCAGGGAAG 1902
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   570  TATGCGTGGGTGGCCGTTATGCACGGAAAAGCAAGAGTTTGAGCGAAGTACCGCTCAAGGAAC 511

Qy  1903  TAGGGTTAGCACCGGGAAATGATGACTTGCCGCCGATGCACTTGCCGCCGATGTGAGTTCTTGAGT 1962
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   510  TAGGTTTAGCACCGGGATGGTCCGCCGCCGATGATCATTTGCCGCCGATGTGAGTTCTTGAGT 451
```

```
FEATURES             Location/Qualifiers
     source          1..855
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="AUIIB10"
                     /clone_lib="AUII"
                     /note="Vector: pHOS2; Site 1: BstXI; 2-3 kb sheared BAC
                     DNA inserted into pHOS2 using BstXI linkers"
ORIGIN
Query Match      23.2%;  Score 725.6; DB 28; Length 855;
Best Local Similarity 91.2%;  Pred. No. 5e-122;
Matches 781;  Conservative 0; Mismatches 74; Indels 1; Gaps 1;

Qy  1494  CCCAGGTTGACAGGGAAGAGCTTAAGATTGTGGGCTACTATTGGGAACAAATCTACCG 1553
          |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   855  CCCAGATTTGACAGAGAAGAGCTTAAGATTGTGGACCACTATAGGGAACAAATCTACCG 796

Qy  1554  CTAAACTCGACGCGGTCCAAGAGCAACCAAATCCGGAGTCCTGTGATTCGCTACTTTCAG 1613
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   795  CTAAACTCGGCGCGGTTTAAGGGCCAACCAAATCCTGTGATCCTGTGATCTCGCTACTTTCAG 736

Qy  1614  CGCTCGGTTGCCAATGTTTTTTACTCCAGGGAGTCTACAGGCACCGTGTCTAACACAGAC 1673
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   735  CGATCGGTCGCGCAACGTTCTTTTACTCCAGGGAGTCTACAGACACCGTGTCTAACACGGAC 676

Qy  1674  ATGAAGATGATAGATCCAGCGCTTATAGGGATTCTCCGCCTTACAGGGATTCTCCGGGGTGTC 1733
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   675  ATGGAGATGATAGATGCGCTTACAGGGATTCTCCGCCGGACGAAGGGGAAGAATGTC 616

Qy  1734  CTGAGAGAGATCTTAACGACTCACCACCACCACCAGTAATGCCTCGTTGATCATCTGTGTGGG 1793
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   615  TTGAGAGGCGATCTCAACAACGCCACCACCACCAGTAATGCCTCTGTTGATCCATCGTGTGTGGA 556

Qy  1794  TACATGAAGTGGGGCGCTGACAAAGGCAAGAGAAGAAGGCACTATGCGTGGGT 1853
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   555  TACAGGAGTGGGCGTTGACGAAGCGGGTTATGAGGGAAGAAGGTGAGAGGAGCACTATGCGTGGGT 496

Qy  1854  GGCGTTGTGACGCCAATTCTGAAAGTTTGTGGAGTTCCGCTCAAGGAAGTAGGGTTAGCA 1913
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   495  GGCGTTATGACGCCCAATTCTGAAGCTTCTGAAGTACCGCTCAAGGAACTAGGTTTAGCA 436

Qy  1914  CCGAGAATGATGACTTGGATCACTTGGCCGATGTGAGTTCTTCTGAGTTTGACATGGTT 1973
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   435  CCGAGGATGATGACTTGGATCATTTGCGCAGATGTGATCATTTGCGCAGATGTGAGTTCTTGGAGTTTGACATGGCT 376

Qy  1974  GGCGACTTTCACCGCTACAGGTTCGAGCATTCATCGATTAGAATCGCCAACATTCTTTTC 2033
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   375  GGTGACTTTCACCGCTACAGGTTCGAGCATTCATCGATTAAAAATCGCCAACATTCTCTC 316

Qy  2034  CCCTGCATTTACGGTTACGCTACTAGGATTCTCGAGGGCGCAGGAACATTGACTTCAAGCCTGCGCTT 2093
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   315  CCCTGCATTTACGGTTACGCTACTAGGATTCTCGAGGGCGCAGGAACATTGACTTCAAGCCTGCGCTT 256

Qy  2094  GAAGATCTTATTTCGAGGGCAGTCCGCCAACTGAGGAGATTAGTCACACCGAAGGAGCT 2153
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   255  GAAGAACTTTATTTCGAGGGCAGTCCGCCAACTGAGGAGATCAGTCAGTCACACCGAAGGAGCT 196

Qy  2154  ACAATAGAAGAGATGTTGATGAGACATATGATAGATGAGGCGGAGTTTGACACGAGCATG 2213
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   195  ACAACAGAAGAGATATTGATGAGACGGATGAGATAGATGAGGCGGAGTTTGACACGAGCATG 136

Qy  2214  TATCATTCAGTGAGCAATTACCTTCCAGGGAGAAAGCAAGAGTTTGAGCGAAGCTCAC 2273
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   135  TATCATTTCAGTGAGCAATATACCTTCCTGCGAGGGAGAAAGCAAGAGTTTGAGCGAAGCTCAC 76

Qy  2274  AGGAACAACAGCAAGCTGCAGAAGTGGTCGCAAGAAACAGGAGATAAGTAGTTACTCGCCAAGTGC 2333
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    75  AGAAACAACAAGCTGTAGAAGTGGTCATGAAACAGACAAGTTACTCGCCAAGT-C 17
```

```
Query Match        21.9%;  Score 682.2;  DB 28;  Length 801;
Best Local Similarity   90.9%;  Pred. No. 4e-114;
Matches  726;  Conservative  0;  Mismatches  73;  Indels  0;  Gaps  0;

Qy  1623 GCCAATGTTTTTTACTCCAGGGAGTCTACAGGCACCGTGCTCTAACACAGACATGAGAATG 1682
         ||||| || | |||||||||||||||||||| ||||||||||||||||| ||||||||||||
Db     2 GCCAACGTCTTTTACTCCAGGGAGTCTACAGACACCGTGCTCTAACACGGACATGAGAATG   61

Qy  1683 ATAGATTCACGCGCTTATAGGGATTCCCGCCTTACAAAAGGAAAGAAATGTCCTGAGAGGA 1742
         ||||||| | ||||||| |||||||||||||    || ||| |||||||||| |||||||
Db    62 ATAGATTATGCGCTTACAGGGATTCCCGCCAACTGAGGGAAGAAGTGTCTTGAGAGGC    121

Qy  1743 GATCTTAAGACTCACCACCAGTAATGCCTCTGTTGATCCATCTGTGTGGTACATGAAG 1802
         |||||  | | ||||||||||||||||||||||||||||||||||||| ||||| |||
Db   122 GATCTCAACAACGACCACCACCAGTAATGCCTCTGTTGATCCATCTGTGTGATACAGGAAG  181

Qy  1803 TGGGCGCTGACAAACGGCAAGAGAAGGTAAGAGGAGCACTATGCGTGGGTGGCCGTTGTG 1862
         |||| | ||||  | ||  |||||| |||| || ||||||||||||||||||||| | ||
Db   182 TGGGGGTTGACGACGGGAAGAGAAGAGGTGAGACGAGCACTATGCGTGGGTGGCCGTTATG  241

Qy  1863 ACGCCAATTCTGAAAGTTTGTGGAGTTCCGCTCAAGGAAGTAGGGTTAGCACCGAGAATG 1922
         |||||||||||  |||| ||||| || |||||||||||| |||| |||||||||||| |||
Db   242 ACGCCAATTCTCGGAAGCTTGTGAAGTACCGCTCAAGGAACTAGGTTTAGCACCGAGGATG  301

Qy  1923 ATGGACTTGGATCACTTGCGCCGATGTGAGTTCTCTGAGTTTGACATGGTTGGGGACTTT 1982
         |||||||||||||| ||||| |||||||||||  |||||||||||||||| ||| ||||||
Db   302 ATGGACTTGGATCATTTGCGCAGATGTGAGTTTGAGTTTGGAGTTTGACATGGCTGGTGACTTT  361

Qy  1983 CACCGCTACAGGTTCGAGCATTCTTTCCCTGCGCTTGAAGATCTT 2042
         |||||||||||||||||||||||  ||||||||||||||||| ||
Db   362 CACCGCTACAGGTTCGAGCATTCAAGCCTGCGCTTGAAGAACTT  421

Qy  2043 TACGCTACTAGGATTCTGAGGGCAGGAACATTGACTTCAAGCCTGCTTGAAGATCTT 2102
         ||||||||||||||||||||||||||||||||||||||||||||| ||||||| ||
Db   422 AACGCTACTAGGATTCTGAGGGCAGGAACATTGACTTCAAGCCTGCTTGAAGAACTT  481

Qy  2103 TATTTCGAGGGCAGTCCGCCAACTGAGGAGATTAGTCACACGGAGGAGCTACAATAGAA 2162
         |||||||||| |||||| |||||||||||||| |||||  |||||||||||||| ||||
Db   482 TATTTCGAGGGTAGTCCCCAACTGAGGAGATCAGTCATACAGGAGGAGCTACAACAGAA  541

Qy  2163 GATGTTGATGAGACATATGATATAGATGAGGCGGAGTTTGACACGAGCATGTATCATTTC 2222
         ||| ||||||||||  ||||||||||||||||| | |||||||||||||||||||||||||
Db   542 GATATTGATGAGACGGATGATATAGATGAGGCGACCGAGTTTGACACGAGCATGTATCATTTC  601

Qy  2223 AGTGAGCATATACCTCCTGCGAGGAAAAGCAAGAGTTTGAGGCGAAGCTCACAGGAACAAC 2282
         |||||||||||||||||||||||| ||| | |||||||||||||||||||||||  |||||
Db   602 AGTGAGCATATACCTCCTGCGAGGGAAAACAGGAGTTTGAGGCGAAGCTCACAGAAACAAC  661

Qy  2283 AGCAAGCTGCAGAAGTGTGCAGAAGATCAGGATGAAGTTACTGCCAAGTGCCTCAGGGCT 2342
         | |||||||  |||||||||| | |||||| ||||| ||| ||||||||||||||||||||
Db   662 AACAAGCTGTAGAAGTGTGCATGAAACAGGACAAGCTTCACTGCCAAGTGCCTCAGGGCT  721

Qy  2343 ATCAAGTTTCTGAAGGACAAGATCAGCTGCTCCTCTTCCACTACAACTATTCCGCAATGA 2402
         ||  |||||  ||||  || ||| | ||| | ||| ||| |||||| |  ||||||  |
Db   722 ATTCAGTTTATGAAGAAACAGCTCTGTCGTCTCCTTCCTCCACTACAGCGGATTCCGCAGGGA  781

Qy  2403 CAGCTCCCTCAGGACATGC 2421
         || ||||||||||||||||
Db   782 CATCTCCCTCAGGACATGC  800
```

```
Qy  1963 TTGACATGGTTGGCGACTTTCACCGCTACACAGGTTCGAGCATTCATCGATTAGAATCGCCA 2022
         ||||||||| ||| |||||||||||||||||||||||||||||||||||||||| ||||||||
Db   450 TTGACATGGCTGGTGACTTTCACCGCTACACAGGTTCGAGCATTCATCGATTAAAATCGCCA  391

Qy  2023 ACATTCTTTTCCCTGCATTTACGCTACTAGGATTCTGAGGGCAGGAACATTGACTTCA 2082
         |||||||| ||||||||||  |||||||||||||||||||||||||||||||||||||
Db   390 ACATTCTTCTCCCTGCATTAACGCTACTAGGATTCTGAGGGCAGGAACATTGACTTCA  331

Qy  2083 AGCCTGCGCTTGAAGATCTTTATTTCGAGGGCAGTCCGCCAACTGAGGAGATTAGTCACA 2142
         ||||||||||||||||  ||||||||||||| ||||| |||||||||||| |||||| |||
Db   330 AGCCTGCGCTTGAAGAACTTTATTTCGAGGGTAGTCCCCAACTGAGGATCAGTCATA  271

Qy  2143 CCGAAGGAGCTACAATAGAAGATGTTGATGAGACATATGATATAGATGAGGCCGAGTTTG 2202
         |||| | |||||||| |||||| | |||||||||  ||||||||||||||||| |||||||
Db   270 CCGAACGAGCTACAACAGAAGATATTGATGAGACGGATGATATAGATGAGGCGACCGGAGTTTG  211

Qy  2203 ACACGAGCATGTATCATTTCAGTGAGCATATACCTCAGCGAGGAAAGCAAGAGTTTGA 2262
         |||||||||||||||||||||||||||||||||||| | |||| |||||||| ||||||
Db   210 ACACGAGCATGTATCATTTCAGTGAGCATATACCTCCTGCGAGGGAAAGCAGGAGTTTGA  151

Qy  2263 GCGAAGGTCACAGGAACAACAGCAAGCTGCAGAAGGTGTGCAGAAGATAAGGTAC 2322
         |||||| |||||| ||||||||  ||||| ||||| |||||| |||| |||| ||
Db   150 GCGAAGCTCACAGAAACAACAACAACAAGCTGTAGAAGTGTGTGATGAAACAGGACAAGTTAC  91

Qy  2323 TCGCCAAGTGCCTCAGGGCTATCAAGTTTCTGAAGGACAAGATCAGCTGCTCCTCTTCCA 2382
         |||||||||||||||||||||||||||| ||||||||||||| | |  | ||||||||||
Db    90 TCGCCAAGTGCCTCAGGGCTATCAAGTTTATGAAGGACAAGAACAAGCCTCGTGCTCCTCTTCCA  31

Qy  2383 CTACAACTATTCCGCAATGACAGCTCCCTC 2412
         ||||| || |||||||| || |||||||||
Db    30 CTACAGCGATTCCGCAAGAACATCTCCCTC   1
```

RESULT 3
BH237501
LOCUS   BH237501   801 bp   DNA   linear   GSS 13-NOV-2001
DEFINITION AUIIB05TR AUII Arabidopsis thaliana genomic clone AUIIB05, genomic
            survey sequence.
ACCESSION   BH237501
VERSION   BH237501.1 GI:16908045
KEYWORDS   GSS.
SOURCE   Arabidopsis thaliana (thale cress)
  ORGANISM Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 801)
  AUTHORS   Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Utterback,T.V.,
            Feldblyum,T.V. and Fraser,C.M.
  TITLE   Survey sequencing of Arabidopsis thaliana BAC T2P24
  JOURNAL   Unpublished (2001)
COMMENT   Other GSSs: AUIIB05TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            From Wash. U contig 849.
            Seq primer: TR
            Class: sheared ends.
FEATURES   Location/Qualifiers
     source   1..801
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="AUIIB05"
                     /clone_lib="AUII"
                     /note="Vector: pHOS2; Site_1: BstXI; 2-3 kb sheared BAC
                     DNA inserted into pHOS2 using BstXI linkers"
ORIGIN

RESULT 4
AQ958311
LOCUS   AQ958311   817 bp   DNA   linear   GSS 28-JAN-2000
DEFINITION LERAW77TR LERA Arabidopsis thaliana genomic clone LERAW77, genomic
            survey sequence.
ACCESSION   AQ958311
VERSION   AQ958311.1 GI:6786012
KEYWORDS   GSS.
SOURCE   Arabidopsis thaliana (thale cress)
  ORGANISM Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

```
Qy    1912  CACCGAGAATGATGACTTGGATCACTTGGCGCCGATGTGAGTTCTCTGAGTTTGACATGG 1971
            ||||||| |||| ||| ||||||||||||||||||||||||||||||||||||||||||||
Db     600  CACCGAGGATGACTTAGATCATTGCCCGATGTGATTTCTTGGAGTTTGACATGG 659

Qy    1972  TTGGCGACTTTCACCGCTACAGGTTCGAGCATTCATCGATTAGAATCGCCAACACATTCTTT 2031
            |||||||| ||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     660  CTGGCGACTTTCACCGCTACAGGTTCGAGCATTCATCGATTAAAATCGCCAACACATTCTTC 719

Qy    2032  TCCCCTGCATTTACGCTACTAGGATTCTCGAGGGCAGGAACATTGACTTCAAGCCTGCGC 2091
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     720  TCCCCTGCATTGATGTCTACTAGGATTCTCGAGGGCCAGGAGGATTGACTTCAAGCCTGCGC 779

Qy    2092  TTGAAGATCTTTATTTCGAGGGCAGTCCGCCAACT 2126
            ||||||||| |||| |||||||||||||||||||||
Db     780  TTGAAGA-ATTTATTTCGAGGGCAGTCCACCACT 813
```

RESULT 5
BH244328
LOCUS   AUJA71TF AUJA Arabidopsis thaliana genomic clone AUJA71, genomic
            survey sequence.   781 bp   DNA   linear   GSS 13-NOV-2001
DEFINITION AUJA71TF AUJA Arabidopsis thaliana genomic clone AUJA71, genomic
            survey sequence.
ACCESSION   BH244328
VERSION   BH244328.1 GI:16921164
KEYWORDS   GSS.
SOURCE   Arabidopsis thaliana (thale cress)
  ORGANISM Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 781)
  AUTHORS   Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Utterback,T.V.,
            Feldblyum,T.V. and Fraser,C.M.
  TITLE   Survey sequencing of Arabidopsis thaliana BAC f4i2
  JOURNAL   Unpublished (2001)
  COMMENT   Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            From Wash. U contig 1258.
            Seq primer: TF
            Class: sheared ends.
FEATURES   Location/Qualifiers
     source   1..781
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="AUJA71"
                     /clone_lib="AUJA"
                     /note="Vector: pHOS2; Site 1: BstXI; 2-3 kb sheared BAC
                     DNA inserted into pHOS2 using BstXI linkers"
ORIGIN

Query Match   21.2%; Score 663.4; DB 28; Length 781;
Best Local Similarity   90.9%; Pred. No. 1.1e-110;
Matches   706; Conservative   0; Mismatches   71; Indels   0; Gaps   0;

```
Qy    1322  AATCGAGTTTCTTTCCACACTGCAAGTGAAATGTATGAGGGACTCACAGACTTTGAGCT 1381
            |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db       5  AATCGAGTTTCTTTCTCCCTGCAAGTGGAGATGTATGAGCGACACTCACGGACTTTGAGCG 64

Qy    1382  GGATACCATGGGTTAGGCTTCTTAGTGCGTTCTTAGTGCGTTACCAGATTTA 1441
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65  GGATACCATGGGGTTAGGATTCTTGACGTTCTCAGTGAATGAACAGCGGTATCAGCTTTC 124

Qy    1442  GATCAAGAAATTGGAAGAACGTGTTTGGTTTCCCTAGTGAAAGGGAACCAACCCCAGGTT 1501
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     125  GATCAAGAAGTTGGAAGAACGTGTTTGGTTTCCCTAGTGAAAGGGAACCAAACCCAGATT 184

Qy    1502  TGACAGGGAAGAGCTTAAGGATTGTGGGCTACTATTGGGACAAATCTACCCGCTAAACTC 1561
```

```
                              /db_xref="taxon:3702"
                              /clone="AUIIB35"
                              /clone_lib="AUII"
                              /note="vector: pHOS2; Site 1: BstXI; 2-3 kb sheared BAC
                               DNA inserted into pHOS2 using BstXI linkers"

ORIGIN

          Query Match      20.7%;  Score 647.2;  DB 28;  Length 848;
          Best Local Similarity   87.5%;  Pred. No. 9.2e-108;
          Matches 740;  Conservative  0;  Mismatches  83;  Indels  23;  Gaps  2;

Qy   2021 CAACATTCTTTCCCCTGCATTTACGCTACTAGGATTCTCGAGGGCAGGAACATTGACTT 2080
          | ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5 CCACATTCTTCTCCCCTGCATTAACGCTACTCGGATTCTCGAGGGCAGGAACATTGACTT 64

Qy   2081 CAAGCCTGCGCTTGAAGATCTTATTTGTGGGTACATGAA 2140
          |||||||||||||||||| ||| |||||||||||||||||
Db     65 CAAGCCTGCGCTTGAAGAACTTTATTTCGAGGGTAGTCCGCCAACTGAGGGAGTCAGTCA 124

Qy   2141 CACCGAGGGAGCTACAATAGAAGATGTTGATGAGACCATATGATATGAGCGCGGAGTT 2200
          | |||||||||||| ||||||||||||||| | | |||||| |||| || |||||||||
Db    125 TACCGAACGAGCTACAACAGAAGATATTGATGAGACGGATGATATAGATGCGACGGAGTT 184

Qy   2201 TGACACGGACATGTATCATTTCAGTGAGCATATACCTCCAGCGAGGAAAAGCAAGAGTTT 2260
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    185 TGACACAGGACATGTATCATTTCAGTGAGCATATACCTCCTGCGAGGAAAAGCAAGGAGTTT 244

Qy   2261 GAGCGAAGCTCACAGGAACAACAGCAAGCGTGTGCAGAAGTGGTGCAGAAACAGGATAAGTT 2320
          |||||||||||||||||||||||||| |||||||||||||||||||||||| ||||||||||
Db    245 GAGCGAAGCTCACAGGAACAACAACAGCGTGTAGAAGTGGTGCATGAAGAAACAGGACAAGTT 304

Qy   2321 ACTCGCCAAGTGCCTCAGGGCTATCAAGTTCTGAAGGACAGAGCAAGATCAGCTGCCTCCTCTTC 2380
          |||||| ||||||||||||||||||| |||| |||||||||||||||||||||||||||||||
Db    305 ACTCGCCAAGTGCTCAGGGGTATCAAGTTTATGAAGACAGAGCAAGATCTCTGCTGCTCCTCTTC 364

Qy   2381 CACTACAACTATTCCGCAATGACAGCTCCCTCAGGAGATAGCAGCACATGCCTTCGAGGAGATATGACGC 2440
          |||||||| ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    365 CACTACAGCGATTCCGGAGGACATTCCCTCAGGAGATAGCAGCACATGCCTTCGAGGAGATATGACGC 424

Qy   2441 GCCCGA-------------------------------GCCTAGAGAGCAGAAGAATTCTGCATGTCCCTGC 2479
          |||||
Db    425 ACCCGAGCCAAGTGCGACGCAGGCCTTTGCCTAGAGAGCAGCAGGAGATTCCGCATGTCCCTC 484

Qy   2480 GAGGCATTCATCATTCGAGCCTCGTGAATCTAGGAGATCAGGAGAAATAGGAGAACGACACTCACTCG 2539
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    485 GAGGCATTCATCATTCAGCCCTCGTGAATCTGGAATCTGGAGGAAGAGGAGGAGGACCACCACTCACTCG 544

Qy   2540 ATCTAGCAGCAGGAGCAGAGACTTCTGCAGTCTCGTAGTTTACGCGACCGGCGTGCTGG 2599
          ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    545 ATCTAGCAGTAGGAGGAGCACCGACTTCTGCAGTCTCGTAGTTTACGCGACCGGCGTGCTGG 604

Qy   2600 CCGCAATAGAAGAAGAGGAGGTCGAGTATCCTCAGAGCGGTGCGCCGCCACAGAGCTGA 2659
          |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    605 CCGCAGCAAGGAGGAGGAGAGGTCGAGTTCCTCAGAGCGGTGGCCGCCGCCACAGAGCTGA 664

Qy   2660 TGAGATCGAGTACCCACCATGCTGGAGCTGGAGACGGAACATGCGCGGTTCGTCTATGGCGTTG 2719
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    665 TGAGGTCGAGTACCCACCCGCTGGAGCTGGAGCTGATGGCAGAACAAGACGGTTCGTCTATGGCCTG 724

Qy   2720 GGAGCAATCACAGGCAGCAGGCCATTGACTACCAACTTCGTTCATTATTGGACTGAGGTAAGCG 2779
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    725 GGAAGCATCGCAGGCAGCAGGCCATTGACGAGCAACTACGTTCTTCGACTGAGGTAAGCG 784

Qy   2780 CCTCACTTCACCATTATATTATATCATCTCTTGTGATTTGTTCTTTATTTTGTTTCAGTG 2839
          |||||||||||||||  ||||| |||| |||||||||||||||| ||||||
Db    785 CCTCACTTCACCAT--GGATATACATCATCTTTTGGGATTGGGTTCTCATTTTTTGGTCTGG 842

Qy   2840 ATTGGA 2845
          ||||||
Db    843 GATGGA 848

RESULT 7
```

```
Db    185 TGACAGGAGAGCTTAAGGATTTGTGGGCCACTATAGGGAACAATCTACCGCTAAACTC 244
          |||||| |||||||||||||||||||||||  |||  |||||||||||||||||||||
Qy   1562 GACGCCGGTCCAAGAGCAACCAAATCCGGAGTCCTGTGATCGGCTACTTTCAGCGCTCCGT 1621
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    245 GGGCGCGGTCTAAGAGCAACCAAATTCGAAGTCCGTGATCCGCTACTTTCAGCGAATCGGT 304

Qy   1622 TGCCAATGTTTTTACTCCAGGGAGTCTACAGGCACCGTGTCTTAACACAGACATGAGAT 1681
          ||||| |||  ||||| ||||||||||||||||||| || ||||||||||||||||||||
Db    305 CGCCACGTCTTTTACTCCAGGGAGTCTACAGGCACCGTGTCTTAACACGGACATGAGAT 364

Qy   1682 GATAGATTCAGCGCTTATAGGGATTCTCCGCCTTACAAAAGGAAAGAATGTCCTCGAGAGG 1741
          ||||||||||||||||||| |||||||||||| |||||||||||||||||||||||||||||
Db    365 GATAGATTATCGCGCTTACAGGGATTCTCCGCCGGACGAAGGGGAAGAAGTGTTCTGAGAGG 424

Qy   1742 AGATCTTAACGACTCACCACCAGTAATGCCTCGTTGATCCATCTGTGTGGGTACATGAA 1801
          |||||| |||||||||||||||| ||||||||||||| ||  ||||||||||||||| |||
Db    425 CGATCTCAACAACGCACCACCAGTAATGCCTCGTTGTTGATCCATCTGTGTGGATACAGGAA 484

Qy   1802 GTGGGGCGCTGACAAACGCCAAGAAGGAGTAAGAGAGCACTACTATGCGTGGGTGGCGTTGT 1861
          ||| |||||||||||||  |||||||||||||||||| |||||| |||||||||||||||||
Db    485 GTGGGGCGTTGACGAACGGGAAGAAGGAGTGACGAGCACTACTATGCCTGGGGTGGCGTTGT 544

Qy   1862 GACGCCAATTCTGAAAGTTTGTGGAGTTCCGCTCAAGGAAGTAGGGTTAGCACCGGAGAAT 1921
          |||||||||||||||||||||||||||||||| || |||||| | ||||||||||||||||
Db    545 GTCGCCAATTCTGGAAGCTTGTGTGAGGAGCTTCAAGGAACTCAGGTTTAGCACCGAGGAT 604

Qy   1922 GATGGACTTGGATCACTTGCCGCCGATGTGAGTTCTCTGAGTTTGCACATGGTTGGCGACTT 1981
          ||||||||| || || |||||||||||| |||||||||| || |||| |||||| |||||||
Db    605 GATGGACTTGGATCATTTGCGCGAGCGATGTGAGTTCTTTGGGAGTTTGCACATGGCGTGGCGACTT 664

Qy   1982 TCACCCGCTACAGGTTCGAGCATTCATCATCGATTAGAATCGCCAACATTCTTTTCCCCTGCAT 2041
          ||||| |||| |||||| ||||||||||| || || |||||||||| |||| |||||||||||||
Db    665 TCACCCTACAGGTTCGAGCATTCATCATCGATTTAAATCGCCCACATTCTCTCCCCTGCAT 724

Qy   2042 TTACGCTACTAGGATTCTCGAGGGCAGGAACATTGACTTCAAGCCTGCGCTTGAAGA 2098
          ||||||||||||||||||||||||||||||||||||||||  |||||||||||||||
Db    725 TGACGCTACTAGGATTCTCGAGGGCAGGAACATTGACTTCAGGCCTGCGCTTGAGA 781

RESULT 6

BH237552

LOCUS       BH237552                 848 bp    DNA     linear   GSS 13-NOV-2001
DEFINITION  AUIIB35TF AUII Arabidopsis thaliana genomic clone AUIIB35, genomic
            survey sequence.
ACCESSION   BH237552
VERSION     BH237552.1  GI:16908149
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 848)
  AUTHORS   Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Utterback,T.V.,
            Feldblyum,T.V. and Fraser,C.M.
  TITLE     Survey sequencing of Arabidopsis thaliana BAC T2P24
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: AUIIB35TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            From Wash. U contig 849.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..848
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
```

```
BH244305/c
LOCUS       BH244305              813 bp    DNA     linear   GSS 13-NOV-2001
DEFINITION  AUJA72TF AUJA Arabidopsis thaliana genomic clone AUJA72, genomic
            survey sequence.
ACCESSION   BH244305
VERSION     BH244305.1  GI:16921118
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 813)
  AUTHORS   Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Utterback,T.V.,
            Feldblyum,T.V. and Fraser,C.M.
  TITLE     Survey sequencing of Arabidopsis thaliana BAC f4i2
  JOURNAL   Unpublished (2001)
  COMMENT   TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            From Wash. U contig 1258.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..813
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="AUJA72"
                     /clone_lib="AUJA"
                     /note="Vector: pHOS2; Site 1: BstXI; 2-3 kb sheared BAC
                     DNA inserted into pHOS2 using BstXI linkers"
ORIGIN
Query Match         19.9%;  Score 622.6;  DB 28;  Pred. No. 2.8e-103;
Best Local Similarity  87.6%;  Length 813;
Matches  705;  Conservative  0;  Mismatches  69;  Indels  31;  Gaps  1;

Qy    978  GTGAGCAACATGTTTACGAGAGCTATAGGATGGATGAATTTGAACGCCTCTGCAGCTTCGACG  1037
           || ||||| |||||||||||||| |||| |||||||||||||||||||||||||||||
Db    805  GAGAGCAACAACATGTTTACCAGAGCTTCAGGGATGAAGAGGGCGATGTCGAGTAGTAGTAGGCT  746

Qy   1038  TAATCAAAGAAGAGCTGAAATCGCTAGAGAAGAGGGCGATGTCGAGTAGTAGTAGTAGAGCT  1097
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    745  TAATCAGAGAGGGCTGAAATCGCCAGAGAAGAGGGCGATGTCAAGTAGTAGTAGTAGAGCT  686

Qy   1098  GATTGATGAGGATATCAAAACTGAGTATGAGCCAGAGTCATGGCGCAAGGAGACGAAGCT  1157
           ||||||||||||| |||||| ||||| ||||||||||||||||||||||||||||||||
Db    685  GATTGATGAGGCTATCGAAAGTGAGTATGAGCCCGAGTCATGGCGCAAGGAGACGAAGCT  626

Qy   1158  ACTGAACAAATCCGACGAGGTTACAGTGGAGGAGTATATCAGATTCTTGAGATGAATGA  1217
           || ||||| ||||||||||||||||||||||||||||||||||||||||||
Db    625  GCTGAACAAACCAGACGAGAGGTTACAGTGGAAGAGTATATCAGGTTCTTGAGATGAATAA  566

Qy   1218  CTTCTGGGAACGAGCTGAGGTATCCCTGATATGAGACTTAGCCCAGTTGGGGTTACTGGAGGA  1277
           |||||||||||||||||||||||| |||||||||||| ||||||||||| ||||||||||
Db    565  TTTCTGGGAACGAGCGCGGGTATCCGTGCTATGAGACTTAGCCCAGCTAGGGTTACTGGAGGA  506

Qy   1278  CGTGCAGCATCTGTTCGAGAAGTGTCATCTG-----------------------  1308
           ||||||||||||||||||||||||||||||
Db    505  CGTGCAGCATCTGTTCGAGAAGTGTCATCTGGAGAAACTCATGTCTTACCCTTACCCTGC  446

Qy   1309  --ATAAGGAGGAGACAATCGAGTTCTTTCCACACTGCAAGTGGAAATGTTATGAGGGACT  1366
             ||||||||||||||||||||||||||||| |||||||||||||| ||||||||||||
Db    445  CTATAAGGAGGAGACAATCGAGTTCTTTCCCACGTTCAAGTGGAGATGATGTATGAGGGCACT  386

Qy   1367  CACAGACTTTGAGCTGGATACCATGGGGTTAGGCTTCTTTGACGTTCGTTCTTAGTGTGATGAACA  1426
           |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    385  CACGGACTTTGAGCGGGGATACCATGGGGTTAGGATTCTTTGACGTTCGTTCTCAGTGAATGAACA  326
```

```
     1427  GCGGTACCAGATTTAGATCAGAATTGAAGAACTGTTTGGTTCCCTAGTGGAAAGGG  1486
           |||||||||||||||| ||||||||| ||||||||||||||||||||||||||||||
     325  GCGGTATCAGCTTTCGATCAGAAGAGTTGGAAGAACTGTTTGGTTCCCTAGTGGAAAGGG  266

     1487  AACCAACCCCAGGTTGACAGGGGAAGAGCTTAAGAGAGCTTAAGGATTTGTGGGCTACTATTGGAACAA  1546
           |||||||||||||||||||||||||||||| ||||||||||||||| |||||||||||||||||
     265  AACCAAACCCAGATTTGACACGAGGGAAGAGCTTAAGGATTTGTGGGCCACTATAGGGAACAA  206

     1547  TCTACCGCTAAACTCGACGCGGTCCAAGAGCAACCAAACCAAATTCCGAGTCCTGATTCGCTA  1606
           |||||||||||||||||||||||||||||||| ||||||| |||||||||||||||||||
     205  TCTACCGCTAAACTCGGCCGGGTCTAAGAGCAACCAAATTCGAGTCCTGTGTGATCCGCTA  146

     1607  CTTTCAGCGCTCGGTTGCCAATGTTTTTTACTCCAGGGAGTCTACAGGCACCGTGTCTAA  1666
           ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
     145  CTTTCAGCAATCGGTCGCCAACGTCTTTTTACTCCAGGGAGTCTACAGGCACCGTGTCTAA  86

     1667  CACAGACATGAAGATGATAGATAGATCAGCGCTTATAGGGATTCCCGCTTACAGGATTCTTAAC  1751
           ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
     85   CACGGACATGAAGATGATAGATAGATTATGCGCTTACAGGATTCCCGCCGGACCGAAGGGGAA  26

     1727  GAATGTTCCTGAGAGAGGAGATCTTAAC  1751
           |||||||||||||||||||||||||||
     25   GAATGTTTGAGAGAGGCGGATCTCAAC  1
```

```
RESULT 8
BH235546
LOCUS       BH235546              806 bp    DNA     linear   GSS 13-NOV-2001
DEFINITION  AUIVA07TF AUIV Arabidopsis thaliana genomic clone AUIVA07, genomic
            survey sequence.
ACCESSION   BH235546
VERSION     BH235546.1  GI:16905904
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 806)
  AUTHORS   Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Utterback,T.V.,
            Feldblyum,T.V. and Fraser,C.M.
  TITLE     Survey sequencing of Arabidopsis thaliana BAC F5K16
  JOURNAL   Unpublished (2001)
  COMMENT   Other_GSSs: AUIVA07TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            From Wash. U contig 1269.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..806
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="AUIVA07"
                     /clone_lib="AUIV"
                     /note="Vector: pHOS2; Site 1: BstXI; 2-3 kb sheared BAC
                     DNA inserted into pHOS2 using BstXI linkers"
ORIGIN
Query Match         19.6%;  Score 611.6;  DB 28;  Length 806;
Best Local Similarity  87.5%;  Pred. No. 2.8e-101;
Matches  705;  Conservative  0;  Mismatches  69;  Indels  32;  Gaps  2;

Qy   1182  AGTGGAGGAGTATATCAGATTCTTGAGATGAATGAACAGCGGTATCCTCGGGGAACGAGGTATCCCTG  1241
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  AGTGGAAGAGTATATCCCGTTCTTTGAGATGAATGAATAATTCTGGGGAACGAGCGCGGGTATCCGTG  60
```

Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
From Wash. U contig 975.
Seq primer: TF
Class: sheared ends.

FEATURES   Location/Qualifiers
     source   1..681
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="AUIJA07"
                     /clone_lib="AUIJ"
                     /note="Vector: pHOS2; Site 1: BstXI; 2-3 kb sheared BAC
                     DNA inserted into pHOS2 using BstXI linkers"

ORIGIN

Query Match   19.4%; Score 604.6; DB 28; Length 681;
Best Local Similarity 93.5%; Pred. No. 5.7e-100;
Matches 631; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 1242 ATATGAGACTTTAGCCCAGTTGGGGTTACTGGAGACGTGCAGCATCTGTTCGAGAAGTG 1301
         ||||||||||||||| |||||| ||| ||||||||||||||||||||||||||||||||
Db   61 CTATGAGACTTTAGCCCAGCTGGGGTTACTGGAGACGTGCAGCATCTGTTCGAGAAGTG 120

Qy 1302 TCATCTG------------------ATAAGGAGGAGACAATCGAGTT 1330
         |||||||| |||||||||||||||||||||
Db   121 TCATCTGGAGAAACTCATGTCTTACCCTGCCTATAAGGAGGAGACAATCGAGTT 180

Qy 1331 TCTTTCCACACTGCAAGTGAAATGTATAGAGGGACTTCACAGACTTTGAGCTGGATACCAT 1390
         ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   181 TCTTTCCACGCTGCAAGTGGAGATGTATGAGGCACTCACGGACTTTGAGCTGGGGATACCAT 240

Qy 1391 GGGGTTAGGCTTCTGACGTTCTTAGTGATGAACAGACGCGGTTAGATCAAGAA 1450
         |||| |||||||||||||| ||||||||||||||||||||||||||||||||||
Db   241 GGGGTTAGGATTCTTGACGTTCTCAGTGAATGAACAGCGTATCAGCTTTCGATCAAGAA 300

Qy 1451 ATTGGAAGAACTGTTTGGTTTCCCTAGTGAAAGGGAACCAACCCCAGGTTTGACAGGGA 1510
         ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GTTGGAAGAACTGTTTGGTTTCCCTAGTGAAAGGGAACCAACCCCAGATTTGACAGAGA 360

Qy 1511 AGAGCTTAAGGAGTTGTGGGCTACTATTGTGGGAACAATCTACCGCTAAACTCGACGCGGTC 1570
         ||||||||||||||||||||| ||| |||||||||||||||||||||||||||||||||
Db   361 AGAGCTTAAGGAGTTGTGGGCCACTATAGGGAACAATCTACCGCTAAACTCGGCGCGGTC 420

Qy 1571 CAAGAGCAACCAAATCCGAGTCCGTGATTCGCTACTTCAGCGCCTCGGTTGCCAATGT 1630
         ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   421 TAAGAGCAACCAAATCGAAGTCCGTGATCCGCTACTTCAGCGCAATCGGTCGCCAACGT 480

Qy 1631 TTTTTACTTCCAGGGAGTTCACAGGCACCCGTGTCTAACACAGACGAAGATGATAGAGTTC 1690
         |||||||| |||||| |||||||| ||||||||||||||||||||||||||||||||||
Db   481 CTTTTACTTCCAGGGAGTTCACAGGCACCCGTGTCTAACACGGACGAAGATGATAGAGTTC 540

Qy 1691 AGCGCTTATAGGGATTCTCCGCCTTACAAAAGGAAAGAATGTTCCTGAGAGAGGAATCTTAA 1750
         ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TGCGCTTACAGGGATTCTCCGCCGCGACGAAGGGAAGAATGTCTTGAGAGGGCCATCTCAA 600

Qy 1751 CGACTCACCACCAGTAATGCCTCTGTTGATCCATCTGTGTGGGTACATGAAGTGGGGCGCT 1810
         || |||||||||||| ||||||||||||||||| |||||||||||||||||||||||||
Db   601 CAACGCACCACCAGTAATGCCTCTGTTGATCCATCTGTGTGGGTACATGAAGTGGGGCGTT 660

Qy 1811 GACAAACGGCAAGAAGAGGTAAGAGGAGCACTATGCGTGGGGTGGTTGTGACGCGCCAAT 1870
         ||| |||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 GACGAACGGGAAGAAGAGGTGAGAGGAGCACTATGCGTGGGGTGGGTTGTGTCGCCCAAT 720

Qy 1871 TCTGAAAGTTTGTGGAGTTCCGCTCAAGGAAGTAGGGTTAGCACCGGAGAATGATGGACTT 1930
         |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 TCTGGAAGCTTGTGGGAGTACCGCCTCAAGG-ACCAGGTTTAGCACCGGAGGATGATGGACTT 779

Qy 1931 GGATCACTTGCGCCGATGTGAGTTCT 1956
         ||||||||||||||||||||||||||
Db   780 GGATCATTGCGCAGATGTGAGTTCT 805

Qy 2000 GCATTCATGATTAGAAATCGCCAACATTCTTTCCCTGCATTACGCTACTAGGATTCT 2059
         ||||||||||||||||| ||||||||||||| |||||| ||||||||||||||||||
Db   681 GCATTCATGATTAGAAATCGCCAACATTCTTCTCCCTTACATTGACGCTACCTACCAAGGATTCT 622

Qy 2060 CGAGGGCAGGAACATTGACTTCAAGCCTGCGCTTGAAGATCTTTATTCGAGGGCCAGTCC 2119
         ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db   621 CGAGGGCAGGAACATTAACTTCAAGCCTGCGCTTGAAGATCTTTATTCGAGGGCCAGTCC 562

Qy 2120 GCCAACTGAGGAGATTAGTCACACGCGGAGGAGCTACAATAGAAGATGTTGATGAGACATA 2179
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   561 GCCAACTGAGGAGATCAGTCACACGCGGAGGAGCTACAACAGAGGATGTTGATGAGACAGA 502

Qy 2180 TGATATAGATGAGGCGGAGTTTGACACGAGCATGTATCATTCAGTGAGCATATACCTCC 2239
         |||||||||||| ||||||||||||||||||||||||||||| |||||||||||||||
Db   501 TGATATAGATGAGGCGGACTTTGACACGAGCATGTATCATTCAGTGAGCATATACCTCC 442

Qy 2240 AGCGAGGAAAAGCAAGAGTTGAGCGAAGCAAGCTCACAGAGAACAACAGCAAGCTGCGAGAAGTG 2299
         || ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   441 AACGAGGGAAGGAAGAGTTGAGCGAAGCAAGCTCACAGAGAACAACAGCAAGCGTGTAGAAGTG 382

Qy 2300 GTGCAAGAAACAGGATAAGTTACTGCGCCAAGTGCCTCAGGGCATCATCAAGTTCTGAAGGA 2359
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   381 GTGCAAGAAACAGGATAAGTTACTGCGCCAAGTGCCTCAGGGCTATCACGTTTCGTGAAGAA 322

Qy 2360 CAAGATCAGCTGCTCCTCTTCCACTACAACTATTCCGCAATGACAGCTCCCTCAGGACAT 2419
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   321 CAAGATCAGCTGCTCCTCTTCCACTACAGCCATTCCGCAGGAGCAGCTCCCTCAGGACAT 262

Qy 2420 GCCTTCGAGGAGGATATGATGACGCGGCCCGAGCCTTAGAGAGCAGAAGATTCTGACGATGTCCCTGC 2479
         ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   261 GCTTTTCGAGGAGGATATGATGACGTGCCTGAGCCTTAGAGAGCCTAGAGAGTTCTGCATGTTCCCGC 202

Qy 2480 GAGGGCATTCATCATTCCGAGCCTCATGAATCTGATAGAGGAAGAATAGGAGGAAGACACTCACTCG 2539
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   201 AAGGGCATTCATCATTCCGAGCCTCATGAATCTGATAGAGGAAGAATAGGAGGAACGACCACTCACTCG 142

Qy 2540 ATCTAGCACGAGCAGAGACGACTTCTGCAGTCTCCAGTCTCCGGGGTCCGGGGTGCTGG 2599
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   141 ATCTAGCACGAGCAGAGACGACTTCTGCAGTCTCCAGTCTCCTACTTTACGGCACCCGGGGTGCTGG 82

Qy 2600 CCGCAATAGAAGAAGAGAGGTCGAGTATCCTCAGAGCCGGTGCTGGCCGCCGCCAGAGCTGA 2659
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81 CCGCAGCAAGAGGAGGAGAGGTCGAGTACCCCTCGCCAGAGCCGGTGCCGCCGCCGCCACACAACTGA 22

Qy 2660 TGAGATCGAGTACCC 2674
         |||||||||||||||
Db   21 TGAGGTGCGAATACCC 7

RESULT 9
BH244756/c
LOCUS   BH244756   681 bp   DNA   linear   GSS 13-NOV-2001
DEFINITION AUIJA07TF AUIJ Arabidopsis thaliana genomic clone AUIJA07, genomic
            survey sequence.
ACCESSION   BH244756
VERSION   BH244756.1 GI:16921934
KEYWORDS   GSS.
SOURCE   Arabidopsis thaliana (thale cress)
  ORGANISM Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 681)
  AUTHORS   Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Utterback,T.V.,
            Feldblyum,T.V. and Fraser,C.M.
  TITLE   Survey sequencing of Arabidopsis thaliana BAC T19J2
  JOURNAL   Unpublished (2001)
COMMENT   Other_GSSs: AUIJA07TR

```
Qy    2572  CTCGTAGTTTACGGACGACCGGTGCTGGCCGCAATAGAGAAGAGAGGTCGAGTATCCTC  2631
            |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421  CTCGTACTTTACGGACCGGCGGGTGCTGGCCGCAGCAGCAGGAGGTCGAGTACCCTC   480

Qy    2632  AGAGCGGTGCTGGCCGCCACCAGAGCTGATGAGATCGATGACTGAGTGCTGGAGCTGATA  2691
            |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db     481  AGAGCGGTGCTGGCCGCCACCAACTGATGAGGTCGAATACCCACCTGCTGGAGCTGATA   540

Qy    2692  CGGAACATGGCGGGTTCGTCTATGGCTTGGGAGCAATCACAGGCAGCCATTGACTACCAAC  2751
            |||||||||||||||||||||||| |||||||||||||||||||||||| ||||||
Db     541  CAGAACAAGGCGGTTCGTCGTCTATGGCCTGAGACGAATCACAGGCAGCCATTAAGCAGCAAC   600

Qy    2752  TTCGTTCATTATTCGACTGACTGAGGTAAGCGCCCTCACTTCACCATTATTATTATTCATCTCTT  2811
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601  TACGTTCATTCTTCGACTGACTGAGGTAAGCGCCCTCACTTCACCATTGTTATTATATCATCTGGT   660

Qy    2812  GTGATTTGTTCTTTATTTTGT  2832
            |||||||||||||||||||||
Db     661  GTGATTTGTTCTTTCATTTTGT   681
```

RESULT 11
B08362
LOCUS   B08362   1101 bp   DNA   linear   GSS 14-MAY-1997
DEFINITION   T24I16-T7.1 TAMU Arabidopsis thaliana genomic clone T24I16, genomic
             survey sequence.
ACCESSION   B08362
VERSION   B08362.1   GI:2089483
KEYWORDS   GSS.
SOURCE   Arabidopsis thaliana (thale cress)
ORGANISM   Arabidopsis thaliana
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
             rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 1101)
AUTHORS   Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
             Ecker,J.
TITLE   BAC End Sequences at ATGC
JOURNAL   Unpublished (1997)
COMMENT   Other GSSs: T24I16-T7, T24I16-Sp6, T24I16-Sp6.1
             Contact: Ecker J.
             Arabidopsis Thaliana Genome Center
             University of Pennsylvania
             Dept. of Biology, University of Pennsylvania, Philadelphia, PA
             19104
             Tel: 215-898-9384
             Fax: 215-898-8780
             Email: jecker@atgenome.bio.upenn.edu
             Seq primer: T7
             Class: BAC ends
             High quality sequence start: 133
             High quality sequence stop: 810.
FEATURES   Location/Qualifiers
     source   1..1101
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="T24I16"
                     /sex="hermaphrodite"
                     /clone_lib="TAMU"
                     /note="Vector: BeloBACII; Site_1: HindIII; Site_2:
                     HindIII; Produced by Rod Wing"
ORIGIN

Query Match   19.2%; Score 598.6; DB 28; Length 1101;
Best Local Similarity 81.9%; Pred. No. 5.8e-99;
Matches 803; Conservative   0; Mismatches 147; Indels 31; Gaps   9;

```
Qy    1873  TGAAAGTTTGTGGAGTTCCGCTCAAGGAAGTAGGTTAGCACCGAGAATGATGGACTTGG  1932
            ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      84  TGCTTGCTTGTGGAGTCCCCCTCGGAGAACCAGGTTCGGACCCCGATGATGGACTTGG   143
```

RESULT 10
BH244796
LOCUS   BH244796   691 bp   DNA   linear   GSS 13-NOV-2001
DEFINITION   AUIJA58TR AUIJ Arabidopsis thaliana genomic clone AUIJA58, genomic
             survey sequence.
ACCESSION   BH244796
VERSION   BH244796.1   GI:16921996
KEYWORDS   GSS.
SOURCE   Arabidopsis thaliana (thale cress)
ORGANISM   Arabidopsis thaliana
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
             rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 691)
AUTHORS   Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Utterback,T.V.,
             Feldblyum,T.V. and Fraser,C.M.
TITLE   Survey sequencing of Arabidopsis thaliana BAC T19J2
JOURNAL   Unpublished (2001)
COMMENT   Other GSSs: AUIJA58TF
             Contact: Chris Town
             TIGR
             9712 Medical Center Drive, Rockville, MD 20850, USA.
             Tel: 301-838-3523
             Fax: 301-838-0208
             Email: cdtown@tigr.org
             From Wash. U contig 975.
             Seq primer: TR
             Class: sheared ends.
FEATURES   Location/Qualifiers
     source   1..691
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="AUIJA58"
                     /clone_lib="AUIJ"
                     /note="Vector: pHOS2; Site_1: BstXI; 2-3 kb sheared BAC
                     DNA inserted into pHOS2 using BstXI linkers"
ORIGIN

Query Match   19.3%; Score 601; DB 28; Length 691;
Best Local Similarity 92.7%; Pred. No. 2.6e-99;
Matches 631; Conservative   0; Mismatches 50; Indels   0; Gaps   0;

```
Qy    2152  CTACAATGAGAAGATGTTGATGAGACATATGATATAGATGAGGCGGAGTTTGACACGAGCA  2211
            |||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db       1  CTACAACAGAAGATGTTGATGAGACAGATGATATAGATGAGGCCGACTTTGACACGAGCA    60

Qy    2212  TGTATCATTTCAGTGAGCATATACCTCCAGCGGAGGAAAAGCAAGAGTTTGAGCGAAGCTC  2271
            ||||||||| ||||||||||||||||||||| |||| ||||||||||||||||||||||
Db      61  TGTATCATTTAAGTGAGCATATACCTCCCAAGCGAGGGAAGGAAGAGTTTGAGCGAAGCTC   120

Qy    2272  ACAGGAACAACACAGCAAGCTGCAGAAGTGGTGCAGAAACAGGATTAAGTTACTCGCCAAGT  2331
            |||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db     121  ACAGGAACAACACAGCAAGCTGTAGAAGTGGTGCAGAAACAGGATAAGTTACTCGCCAAGT   180

Qy    2332  GCCTCAGGGCTATCAAGTTTCTGAAGGACAAGAGCAGCTGCTCCTTCCACTACAACTA  2391
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181  GCCTCAGGGCTATCACGTTTCTGAAGGACAAGAGCAGCTGCTCCTTCCCACTACAGCTA   240

Qy    2392  TTCCGCAATGACAGCTCCCTCAGGACATGCCTTCGAGGAGATATGACGCGCCCGAGCCTA  2451
            ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db     241  TTCCGCAGGAGCAGCTCCCTCAGGACATGCCTTTGCGAGGAGATATGACGTGCCTGAGCCTA   300

Qy    2452  GAGAGCAGAAGAGATTCTGCATGTCCCTGCGAGGCATTCATCATTCGAGCCTCGTGAATCTA  2511
            |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     301  GAGAGTAGGAGAGATTCTGCATGTTCCGCAAGGCATTCATCATTCGAGCCTCATGAATCTA   360

Qy    2512  GGAAGAATAGGAGAACGACACTCACTCGATCGATCTAGCAGGAGCAGAGACTTCTGCAGT  2571
            |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db     361  GGAGGAAGAAGAGGAGCCACACTCACTCGATCGATCTAGCAGGAGCAGAGACTTCTGCAGT   420
```

SOURCE   Arabidopsis thaliana (thale cress)
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 704)
AUTHORS   Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Utterback,T.V.,
            Feldblyum,T.V. and Fraser,C.M.
TITLE   Survey sequencing of Arabidopsis thaliana BAC T2P24
JOURNAL   Unpublished (2001)
COMMENT   Other_GSSs: AUIIA17TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            From Wash. U contig 849.
            Seq primer: TF
            Class: sheared ends.
FEATURES   Location/Qualifiers
     source   1..704
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="AUIIA17"
                     /clone_lib="AUII"
                     /note="Vector: pHOS2; Site_1: BstXI; 2-3 kb sheared BAC
                     DNA inserted into pHOS2 using BstXI linkers"
ORIGIN

Query Match   18.8%; Score 587.4; DB 28; Length 704;
Best Local Similarity 90.5%; Pred. No. 7.6e-97;
Matches 638;   Conservative   0; Mismatches   66; Indels   1; Gaps   1;

Qy   1933 ATCACTTGCGCCGATGTGAGTTCTCTGAGTTTGACATGGTTGGCGACTTTCACCGCTACA 1992
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   144 ATCATTTGCCAGATGTGAGTTCTTGGAGTTTGACATGGCTGGCGACTTTCACCGCTACT 203

Qy   1459 AACTGTTTGGTTTCCCTAGTGGAAGGGAACCACCCCAGGTTGACAGGGAGGAAGAGCTTA 1518
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   704 AACTGTTTGGTTTCCCTTGTGAAAAGGGAACCAAACCCAGATTTGACAGAGAGGAGCTTA 645

Qy   1993 GGTTCGGAGCATTCATCGATTAGATAGAATCGCCAACATTCTTTCCCCTGCATTACGCTACTA 2052
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   204 GGTTCGGAGCATTCATCGATTAGATAGAATCGCCAACATTCTCTCCCCTGCATTGACGCTACTA 263

Qy   1519 AGGATTTGTGGGCTACTATTGGGAACAATCTACCGTCAAACTCCACGCGGTCCAAGAGCA 1578
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   644 AGGATTTGTGGGCTACCACTATAGGGAACAATCTACCGTCAAACTCCACGCGGT-TAAGGGCA 586

Qy   2053 GGATTCTCGAGGGCCAGGAACATTGACTTCAAGCCTGCGCTTGAAGATCTTTATTTCGAGG 2112
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   264 GGATTCTTGAGGGCCAGGAACATTGACTTCAAGCCTGCGCTTGAAGATCTTTATTTCGAGG 323

Qy   1579 ACCAAATCCGGAGTCCTCGTGATTCGCTACTTCAGCGCGTCGGTTGCCAATGTTTTTTACT 1638
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   585 ACCAAATCCGGAGTCCTCGTGATTCGCTACTTCAGCGCGTCGGTTGCCAATGTTTTTTACT 526

Qy   2113 GCAGTCCGCCAACTGAGGAGATTAGTCACACCGAAGGAGCTACAATAGAAGATGTTGATG 2172
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   324 GCAGTCCGCCAACTGAGGAGTTTAGTCACACCGAAGGAGCTACAACAGAATATGTTGATG 383

Qy   1639 CCAGGGAGTCTACAGGCACCGTGTCTACACAGCACCGTGTCTACACAGATGTCTGAGAGTATGAGATTCAGCGCTTA 1698
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   525 CCAGGGAGTCTACAGAGCACCGTGTCTACACACCGATGTCTACACACGACACATGGAGATCGATAGATTATGCGCTTA 466

Qy   2173 AGACATATCATATAGATGAGGCGGAGTTTGACACGAGCATGTATCATTTCAGTGAGCATA 2232
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   384 AGACGGATGATATAGATGAGGCGGAGTTTGACTCTATCATGTATCATTTCAGTGAGCATA 443

Qy   1699 TAGGGATTCTCCGCCTTACAAAAGGAAAGAATGTCTGAGAGGAATCTTAACGACTCAC 1758
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   465 CAGGGATTCTCCGCGACGAGGGGAACCATGCGTGTTGAGAGGGCGATCTCAACAACGCAC 406

Qy   2233 TACCTCCAGCGAGGAAAAGCAAGAGTTTGAGCGAAGCTCACAGGAACACACAGCAGCTGC 2292
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   444 TACCTCCTGCGAGGGAAAGCAAGAGTTTGAGCGAAGCTCACAGGAACAACAACAGCAAGCTGC 503

Qy   1759 CACCAGTAATGCCTCTGTTGATCCATCTGTGTGGGTACATGAAGTGGGCGCTGACAAACG 1818
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   405 CACCAGTAATGCCTCTGTTGATCCATCTGTGTGGGTGGATACAGGAGTGGGCGTTGACGAACG 346

Qy   2293 AGAAGTGGTGCAAGAAACAGGATAAGTTACTCGCCAGTGCCTCAGGGCTATCAAGTTTC 2352
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   504 AGAAGTGGTGCAAGAAACAGGACAAGTTACTCGCCAGTGCTTAAGGGCTATCAAGTTTC 563

Qy   1819 GCAAGAAGAAGTAAGAGAGCAACTATGCGTGGTGGTGCGTTGTGACGCCAATTCTGAAAG 1878
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   345 GGAAGAAGAAGTGAGAGGAGCACTATGCGTGGTGCGTTATGACGCCAATTCTGAAAG 286

Qy   2353 TGAAGGACAAGATCAGCTGCTCCTTCTTCCACTACAACTATTCCGCAATGACAGCTCCCTC 2412
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   564 TGAAGGACAAGCTTCAGCTGCTCCTTCTTCCACTACAGCTATTCCGCAGTGCTATCAAGTTTC 623

Qy   1879 TTTGTGGAGTTCCGCTCAAGAAGTTCAGGGAAGGTTAGCACCGGAAGAATGATCGGATCACT 1938
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   285 CTTGTGAAGTACCGCTCAAGAAGTTTTAGCACCGGAGGAACTTAGCACCGGACTTGGATCATT 226

Qy   2413 AGGACATGCCTTCGAGGAGATATGACCGCGCCCAG------------CCTA 2451
          ||||||||||||||||||||||||||||||||||||
Db   624 AGGACATGCCTTCGAGGAGATATGACCGCGCCCAGCGGCGCAAGTCGGCGCAGGCCTGATCCTA 683

Qy   1939 TGCGCCGATGTGAGTTCTCTGAGTTTGACATGGTTGGCGACTTTCACCGCTACAGGTTCG 1998
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   225 TGCGCAGATGTGAGTTCTTGGAGTTTGACATGGCTGGTGTGACTTTCACCGCTACAGGTTCG 166

Qy   2452 GAGAGCAGAAGAGATTCTGCATGTCCCTGCGAGGCATTCATCATTCGAGCCTCGTTGAATCTA 2511
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   684 AAGAGCAGGAGATTCTGCATGTCCCTGCGAGGCATTCATCATTCGAGCCTCCTGAATCTG 743

Qy   1999 AGCATTCATCGATTAGAATCGCCAACATTCTTTCCCCTGCATTACGCTACTACTAGGATTC 2058

Qy   2512 GGAAGAAT-AGGAGGAACGAACACTCACTCGATCTAGCAGCA--GGAGCAGACACTTCTGCA 2569
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   744 GGGAGGAGGATGAAGATGAGATCNCNCTCGATCTAACNNTAGGGAGCGGACGACTTCTGCC 803

Qy   2570 GTCTCGTAGTTTACGCGACCGC-GGTGCTGGCCGCA-ATAGAAGAGAGAGAGCTGCGAGTAT 2627
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   804 GTCTCGTAGTTTACGCCAACGCGGGTGCTGCGCCGCCAGCGTCAAAGAGAAGAAGTTCAGTTT 863

Qy   2628 CCTCAGAGCCGGTGCTGCCGCCACAGAGCTGATG-AGATCGAGTACCCACATGCT-GGAG 2685
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   864 CCCCAAGCGGCGGGGCCGCCGCCCAAACTGAAGAGAGGTCNAAATTCCCACCCCGGCTGGGAG 923

Qy   2686 CTGAT--ACGGAACACTAGCTGGTTCGTCGTCTATGGCTTATGCCTGGAGCAATCACAGGGAGCCATTGA 2743
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   924 CTGATCCCACAAACCGGGCGTCTGCCTTGGGCGGGAACCATCGCCGGGCGCCCTTGGA 983

Qy   2744 CTACCAACTTCGTTCATT--ATTCGACTGAGGTAAGCGCCTCACTTCACCATTATTATTAT 2801
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   984 GAAACCACTTCCTTTCTTCCTTCCGAAGGAGGTTGGCCCCCCCCTCCCCTTCCCCTTTTTTTTTTTTT 1043

Qy   2802 ATCATCTCTTGTGATTTGTTC 2822
          |||||||||||||||||
Db   1044 CCCGCGCGGCGTTATTGGTCC 1064

RESULT 12
BH237544/c
LOCUS   BH237544   704 bp   DNA   linear   GSS 13-NOV-2001
DEFINITION AUIIA17TF AUII Arabidopsis thaliana genomic clone AUIIA17, genomic
            survey sequence.
ACCESSION   BH237544
VERSION   BH237544.1 GI:16908128
KEYWORDS   GSS.

```
Db      165  AGCATTCATCGATTAAAAATCGCCAACATTCTTCCCCCTGCATTAACGCTACTAGGATTC   106
                  ||||||||||  ||||||||||||||||| |||||||  ||||||||||||  |||||
Qy     2059  TCGAGGGCCAGGAACATTGACTTCAAGCCTGCGCTTGAGAGATCTTTATTTCGAGGGCAGTC   2118
             |||||||||||||||||||||||||||||||||||||| ||  |||||||||||||| ||||
Db      105  TCGAGGGCCAGGAACATTGACTTCAAGCCTGCGCTTGAAGAACTTTATTTCGAGGGTAGTC    46

Qy     2119  CGCCAACTGAGGAGATTAGTCACACCGAAGGAGCTACAATAGAAG   2163
             |||||||||||||||| |||||| |||  ||  |||||||  ||||
Db       45  CGCCAACTGAGGAGATCAGTCATAGAAAACGAGCTACACAGAGAAG     1
```

RESULT 13
BH237151/c
LOCUS   BH237151   792 bp   DNA   linear   GSS 13-NOV-2001
DEFINITION AUIIB06TR AUII Arabidopsis thaliana genomic clone AUIIB06, genomic
            survey sequence.
ACCESSION   BH237151
VERSION   BH237151.1 GI:16907522
KEYWORDS   GSS.
SOURCE   Arabidopsis thaliana (thale cress)
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 792)
AUTHORS   Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Utterback,T.V.,
            Feldblyum,T.V. and Fraser,C.M.
TITLE   Survey sequencing of Arabidopsis thaliana BAC T2P24
JOURNAL   Unpublished (2001)
COMMENT   Other GSSs: AUIIB06TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            From Wash. U contig 849.
            Seq primer: TR
            Class: sheared ends.
FEATURES   Location/Qualifiers
     source   1..792
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="AUIIB06"
                     /clone_lib="AUII"
                     /note="Vector: pHOS2; Site_1: BstXI; 2-3 kb sheared BAC
                     DNA inserted into pHOS2 using BstXI linkers"
ORIGIN

Query Match   18.8%; Score 585.6; DB 28; Length 792;
Best Local Similarity   87.1%; Pred. No. 1.5e-96;
Matches   679; Conservative   0; Mismatches   69; Indels   32; Gaps   2;

```
Db      539  ATCAGCTTTCGATCAAGAAGTTGGAGGAGAACTGTTGGTTTCCCTTGTGGAAAGGAGAACCA   480
             ||||  |||  |||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     1492  ACCCCAGGTTTGACAGGGAAGAGAGCTTAAGAGATTTCGTTGGGCTACTATTGGGAACAATCTAC   1551
             |||||||| ||||||| |||||||||||||||||| |||||| ||||||||| |||||||||||
Db      479  AACCCAGATTTGACAGAGAAGAGAGCTTAAGAGATTTGTTGGACCACTATAGGGAACAATCTAC   420

Qy     1552  CGCTAAACTGCGACGCGGTCCAAGAGCAACCAAAATCCGGAGTCCGTGATTCGCTACTTTC   1611
             |||||||||||||||||||  |||||||||||||||||| |||||| |||| ||||||||||
Db      419  CGCTAAACTGCGGCGCGGTTAAGGGCAACCAAAATCCGAAGTCCTGTGATCTGCTACTTTC   360

Qy     1612  AGGGCTCGGTGCCAATGTTTTTTACTCCAGGAGTCTACAGGCACCGTGTCTAACACACAG   1671
             |||| |||||||| |||| ||||||||||||||||||||| ||| | ||||||||||||  
Db      359  AGGGATCGGTGCGCAACGTCTTTACTCCAGGAGTCTACAGACACACGTGTCTAACACGG   300

Qy     1672  ACATGAAGATGATAGATTCAGCGCTTATAGGGATTCTCCGCCTTACAAAAGGAAGAAATG   1731
             ||||| |||| |||||||| |||||||| ||||||||||||||  ||  |||| |||||||
Db      299  ACATGGAGATAGATTATGCGCTTACAGGGATTCTCCGCCGGACCGAAGGGGAAGAAATG   240

Qy     1732  TCCTGAGGAGGAGGATCTTAACGACTCACCACCAGTAATGCCTCGTGTTGATCCATCTGTGTG   1791
             || ||||| ||||| |||||||||| ||||||||||||||||||||||||||||||||||||||
Db      239  TCTTGAGGGGGATCTCAACAACGCACCACCAGTAATGCCTCGTGTTGATCCATCTGTGTG   180

Qy     1792  GGTACATGAAGTGGGCCGCTGACAAACGGCAAGAGAAGGTAAGAGAGCCACTATGCGCTGG   1851
             | ||ACA GAAGTGGGCG TGACGAACGGCAAGAAGAAGGTGAGAGGAGCACTATGCGGTGG
Db      179  GATACACGAAGTGGGCGTTGACGAACGGCAAGAAGAAGGTGAGAGGAGCACTATGCGGTGG   120

Qy     1852  GTGGCGGTTGTGACGCCAATTCTGAAAGTTTGTGGAGTTCCGCTCAAGGAAGTAGGGGTTAG   1911
             ||||||||| |||||||||||||||||| |||| |||  |||| |||||| |||||  |||||
Db      119  GTGGCGGTTATGACGCCAATTCTGAAAGCTTGTGAAGCTTGTGTAGTACCGGCTCAAGGAACTAGGTTTAG    60

Qy     1912  CACCGAGAATGATGGACTTGGATCACTTGCGGCCGATGTGAGTTCTCTGAGTTTGACATGG   1971
             ||||||| |||||||||||||||||| |||||  ||||-||||||| ||| |||||||||||
Db       59  CACCGAGGATGATGGACTTGGATCATTTGCGCAGAT-AGAGTTCTTGACTTTGACATGG     1
```

RESULT 14
BH237213
LOCUS   AUIIB28TR AUII Arabidopsis thaliana genomic clone AUIIB28, genomic
                                     670 bp   DNA   linear   GSS 13-NOV-2001
            survey sequence.
ACCESSION   BH237213
VERSION   BH237213.1 GI:16907584
KEYWORDS   GSS.
SOURCE   Arabidopsis thaliana (thale cress)
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 670)
AUTHORS   Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Utterback,T.V.,
            Feldblyum,T.V. and Fraser,C.M.
TITLE   Survey sequencing of Arabidopsis thaliana BAC T2P24
JOURNAL   Unpublished (2001)
COMMENT   Other GSSs: AUIIB28TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            From Wash. U contig 849.
            Seq primer: TR
            Class: sheared ends.
FEATURES   Location/Qualifiers
     source   1..670
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="AUIIB28"
                     /clone_lib="AUII"
                     /note="Vector: pHOS2; Site_1: BstXI; 2-3 kb sheared BAC

```
Qy     1223  GGGGAACCAGGTATCCCTGATATGAGACTTTAGCCCAGTTGGGGTTACTGGGAGGACGTGC   1282
             |||||||| ||||||| ||| ||||||||||||||||||||||||||||||||||||||||
Db      779  GGGGAACGCGGTATCCGTGCTATGAGACTTTAGCCCAGTTGGGGTTACTGGGAGGACGTGC   720

Qy     1283  AGCATCTGTTCGAGAAGTGTCATCTG------------------------------ATA   1311
             |||||||||||||||||||||||||||                            |||
Db      719  AGCATCTGTTCGAGAAGTGTCATCTGGAGACACTGGAGCACACTCCTTACCCTTACCCTGCCTATA   660

Qy     1312  AGGAGGAGGACAATCGAGTTTCTTTCCACACTGCAAGTGGAAATGTATGAGGGGACTCACAG   1371
             |||||||||| ||||||||||||||||| | |||||||||| ||||||||||||  ||||| 
Db      659  AGGAGGAGGAGCAATCGAGTTTCTTTCCACCGCCGCAAGTGGAGATGTATGAGGGCCACTCACCG   600

Qy     1372  ACTTTGAGCTGGATACCACCATGGGGTTAGGCTTCGACGTTCTTAGTGGATGAACAGCGGT   1431
             |||||| |||||||||| ||||||||||||||||||||||||||| |||| |||||||||||
Db      599  ACTTTGGGCTGGATACCATGGGGTTAGGCTTCGACGTTCTTCAGTGAATGAACAGCGGT   540

Qy     1432  ACCAGATTTAGATCAAGAAAATTGGAGAACTGTGTTTGGTTTCCCTAGTGGAAAGGAGGAACCA   1491
```

ORIGIN   DNA inserted into pHOS2 using BstXI linkers"

Query Match   18.5%;   Score 577.2; DB 28; Length 670;
Best Local Similarity 91.3%;   Pred. No. 5.6e-95;
Matches 612; Conservative   0; Mismatches 58; Indels   0; Gaps   0;

Qy   1612 AGGCGTCCGTTGCCAATGTTTTTACTCCAGGGAGTCACAGGCACCGTGTCTAACACAG 1671
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 AGGCGATCGGTCGCCAACGTCTTTTACTCCAGGAGTCACAGACACCGTGTCTAACACG 60

Qy   1672 ACATGAAGATGATAGATTCAGCGCTTATAGGGATTCTCCGCTTACAAAAGAGAAGAATG 1731
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACATGAAGATGATAGATTATGCCGTTACAGGGATTCTCCGCCCGACGAAGGGGAAGAATG 120

Qy   1732 TCCTGAGAGGAGATCTTAAACGACTCACCACCACCAGTAATGCCTCTGTTGATCCATCTGTGTG 1791
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TCTTGAGAGGGCGATCTCAACAACGACCACCAGTAATGCCTCTGTTGATCCATCTGTGTG 180

Qy   1792 GGTACACATGGGCGCGCTGACAAACGCCAAGAAGAGAAGTAAGACGCACTATGCGTGG 1851
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GATACACAGAGAGTGGGCGTTGACGAACGGGAAGAAGAGGTGAGGAGCACTATGCGTGG 240

Qy   1852 GTGGCGGTTGTGTGACGCCCAATTCTGAAAGTTGTGGAGTTCCGCTCAAGAGAAGTAGGGTTAG 1911
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GTGGCGTTATGACGCCCAATTCTGGAAGCTTGTGAAGTACCGCTCAAGAGAACTAGGTTTAG 300

Qy   1912 CACCGAGAATGATGGACTTGGATCACTTGCGCCGATGTGAGTTCTTGAGTTTGACATGG 1971
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db   301 CACCGAGGATGATGGACTTGGATCATTTTGCGCAGATGTGAGTTTCTTGAGTTTGACATGG 360

Qy   1972 TTGGCGACTTTCACCGCTACAGGTTCGAGCATTCATCGATTAGAAATCGCCAACATTCTTT 2031
          ||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   361 CTGGTGACTTTCACCGCTACAGGTTCGAGCATTCATCGATTAAAATCGCCAACATTCTTC 420

Qy   2032 TCCCCTGCATTTACGCTACTAGGATTCTGAGGGCAGGAACATTGACTTCAGCCTGCGC 2091
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 TCCCCTGCATTTAACGCTACTAGGATTCTCGAGGGCACGGAACATTGACTTCAAGCCTGCGC 480

Qy   2092 TTGAAGATCTTTATTTCGAGGGCAGTCCGCCAACTGAGGAGTTAGTCACACCGAAGGAG 2151
          |||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   481 TTGAAGAACTTTATTTCGAGGGTAGTCCGCCAACTGAGGAGATCAGTCATACCGAACGAG 540

Qy   2152 CTACAATAGAGAGTGTTGATGAGACATATGATATGATAGAGGAGATTCACACGAGACA 2211
          |||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   541 CTACAACAGAGATATTGATGAGACGGATGATATAGATGAGCCGAGTTTGACACAGACA 600

Qy   2212 TGTATCATTTCAGTGAGCATTACCTCCAGCGGAGGAAAAGCAAGAGTTTGAGCGGAAGCTC 2271
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 TGTATCATTTCAGTGAGCATTACCTCCTCGCGGAGGAAAAGCGAGGGAGAGTTTGAGCGGAAGCTC 660

Qy   2272 ACAGGAACAA 2281
          |||||||||
Db   661 ACAGAAACAA 670

RESULT 15
BH244293
LOCUS   BH244293   756 bp   DNA   linear   GSS 13-NOV-2001
DEFINITION AUJAA06TF AUJA Arabidopsis thaliana genomic clone AUJAA06, genomic
            survey sequence.
ACCESSION   BH244293
VERSION   BH244293.1   GI:16921094
KEYWORDS   GSS.
SOURCE   Arabidopsis thaliana (thale cress)
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 756)
AUTHORS   Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Utterback,T.V.,
            Feldblyum,T.V. and Fraser,C.M.
TITLE   Survey sequencing of Arabidopsis thaliana BAC f4i2

JOURNAL   Unpublished (2001)
COMMENT   Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            From Wash. U contig 1258.
            Seq primer: TF
            Class: sheared ends.

FEATURES   Location/Qualifiers
     source   1..756
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="AUJAA06"
                     /clone_lib="AUJA"
                     /note="Vector: pHOS2; Site 1: BstXI; 2-3 kb sheared BAC
                     DNA inserted into pHOS2 using BstXI linkers"

ORIGIN

Query Match   18.5%;   Score 576.2; DB 28; Length 756;
Best Local Similarity 88.5%;   Pred. No. 8.1e-95;
Matches 668; Conservative   0; Mismatches 63; Indels   24; Gaps   3;

Qy   2237 TCCAGCGAGGAAAAGCAAGAGTTTGAGCGGAAGCTCACAGGAACAACAGCAGGTGCAGAA 2296
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 TCCTGCGGAGGGGAAGCAGGAGTTCGAGCGGAAGCTCACAGAACAACACAGCAGCTGCAGAA 61

Qy   2297 GTGGTGCAAGAAACAGGATAAGTTACTCGCCAAGTGCCTCAGGCGTATCAAGTTTCTGAA 2356
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 GTGGTGCATGAAACAGGACAAGTTACTCGCCAAGTGCCTCAGGCGTATCAAGTTTATGAA 121

Qy   2357 GGACAAGATCAGCTGCTCCTCTTCCACTACAACTATTCCCGACTACAACGAGCTCCCTCAGGA 2416
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 GAACAAGCTCAGCTGCTCCTCTTCCACTATAGTTATTCCGCAGGGACATCTCCCTCCTCAGGA 181

Qy   2417 CATGCCTTCGAGGAGATATGACGGCGCC------------------GAGCCTAGAGA 2455
          |||||||||||||||||||||||||||| |||||||||
Db   182 CATGCCTTCAAGGAGATATGACGGCGCCCCGGAGCCAAGTCGGCGGCCAAGGGCCTGAGCCTAGAGA 241

Qy   2456 GCAGAAGATTCTGCATGTCCCTGCGAGGCATTCAGCAGCAGCAGACGACTTCTAGGAA 2515
          ||| ||||||||||||||||||||||||||||||||||||||||||| |||
Db   242 GCAGGTGATTCCCATGTCCCATGCGAGGCATTCATCGAGCCTCGAGCCTCGTAAATCGGGAG 301

Qy   2516 GAATAGGAGACGACACTCACTCGATCTAGCAGCAGCAGCGGAGGAAGGAGAGGTCAGTATCCTCAGAG 2575
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   302 GAAGAGGAGGACCACACTCTCTCGATCTAGCAGCAGTAGGAGGCGGACGACTTCGCAGTCTCG 361

Qy   2576 TAGTTTACGGCACCGGCGGTGCCGGCCCGATGAGATCGAGTATCCTCAGAG 2635
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   362 TAGTTTACGGCACCGGCGGTGCCGGCCCGCAGCAGGGAGTACCCACCTGGTGGAGCTGATACGGA 421

Qy   2636 CGGTGCTGGCCGCCCCACAGAGCTGATGAGATCGAGCTGGAGCTGATACGGA 2695
          |||||||||||||||||||||||||||||||||||||||||||| ||||
Db   422 CGGTGTTGGCCGCCCCACAGAGCTGTGATGAGGTGAGCTGACTACCAACTTCG 481

Qy   2696 ACATGGCGGTTCGTCATGGCTTGGGAGCAATCACAGGCAGCCATTGACTACCAACTTCG 2755
          ||||||||||||||||||||||||||||||||||| |||||||||||
Db   482 ACAAGGCAGTTCGTCATGGCCTGGGAGCAATGGCAGCCATTGACGAGCAGCAACTACG 541

Qy   2756 TTCATTATTCGACTGGAGGTAGCGCCTCACTTCACCATTATTATATCATTCTCTTGTGA 2815
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   542 TTCATTCTTCGACTGGAGGTAAGCACCTCACTTCACCATTTATTATTACCATTCTGTCTTGA 601

Qy   2816 TTTG-TTCTTTATTTTGTTTCAGTGATTGGATTTTGTCCTGAGTACTCCTTCCAAGTTTA 2874
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   602 TTTGTTTCTTCATTTTATTTTTGTTGTATTGGATTTTGTCCTGAGTACTCCTTCCAAGTTTA 661

Qy   2875 TTCACACAGTGGACTGTGTGATTTAAGTTTGGGGAGGGGCTCAGGAA--GTATGTTGCAT 2932
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   662 TTCGCACAGTGGACTGTGTGATTTAAGTTTGGGGAGGGGCTCAGGAAGTGTTGGTTGTTGCAT 721

us-10-088-384a-27.rst

Qy   2933 TGTATATATATTTTTAAGTCTGCATTCATCTAAGGCA 2967
            ||| ||||||||| ||||||||||||||||||||||||
Db   722 TGTGTATATTTTTGAGTCTGCATTCATCTAAGGCA 756

Search completed: September 9, 2004, 09:12:58
Job time : 7048 secs